

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 24, 2005, 01:32:23 ; Search time 165 Seconds
(without alignments)
1036.049 Million cell updates/sec

Title: US-10-016-768A-8

Perfect score: 2250

Sequence: 1 MKQMRQFAIEYISKSGTKQ.....GLYNTDSTGSCNKSQKPV 442

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2250	100.0	442	5	Aae24371 Human E93
2	1913.5	85.0	630	4	Abg17942 Novel hum
3	1754	78.0	356	5	Adh48740 NOV10 pro
4	1411	62.7	277	8	Abg60261 Human gen
5	1090	48.4	213	6	Ada54381 Human pro
6	502	22.3	311	7	Adm04876 Human pro
7	499.5	22.2	433	8	Adn35086 Nuclear r
8	499.5	22.2	572	7	Adel1284 Novel hum
9	499.5	22.2	572	7	Adj69960 Human hea
10	294.5	13.1	104	5	Abp32451 Human ORF
11	273	12.1	53	5	Aae24592 Human E93
12	233.5	10.4	54	5	Aae24593 Fish E93
13	229	10.2	53	5	Aae24594 Mouse E93
14	200.5	8.9	1140	4	Abb71145 Drosophil
15	200.5	8.9	1165	5	Aae24372 Fruit fly
16	165	7.3	53	5	Aae24370 Fruit fly
17	163	7.2	53	5	Aae24595 Nematode
18	133.5	5.9	1109	8	Adr70231 Rabbit GM
19	133.5	5.9	3616	8	Adf45530 Chicken A
20	132.5	5.9	848	5	Abg69087 Botulinum
21	132.5	5.9	849	8	Adn11042 Clostridi
22	131.5	5.8	870	8	Abg59193 Human gen
23	131.5	5.8	2288	8	Adk60224 Angiogene
24	131.5	5.8	2288	8	Adk60525 Angiogene
25	131.5	5.8	2288	8	Adp73148 Angiogene

26	131.5	5.8	2375	7	ADJ68471	Adj68471 Human hea
27	131.5	5.8	2492	8	ADK60200	Adk60200 Angiogene
28	131.5	5.8	2492	8	ADK60501	Adk60501 Angiogene
29	131.5	5.8	2492	8	ADP73124	Adp73124 Angiogene
30	131.5	5.8	2492	8	ADQ89534	Adq89534 Human ATR
31	130	5.8	3248	2	AAR99795	Aar99795 Kinetoch
32	127	5.6	533	5	AAU82977	Aau82977 S. cerevi
33	127	5.6	534	6	ABR53199	Abr53199 Protein s
34	127	5.6	534	7	ADK63276	Adk63276 Disease c
35	126.5	5.6	750	8	ADL05858	Adl05858 M. catarr
36	126	5.6	433	8	ADN19824	Adn19824 Bacterial
37	125.5	5.6	2285	4	ABG63057	Abg63057 Drosophil
38	125	5.6	2759	6	AAO16418	Aao16418 Human nuc
39	124.5	5.5	972	5	AAU76761	Aau76761 Plasmodiu
40	124.5	5.5	2954	2	AAAY01632	Aay01632 Amino aci
41	124.5	5.5	3187	7	ADK56031	Adk56031 Rat Prote
42	124.5	5.5	3187	7	ADK56035	Adk56035 Rat Prote
43	123.5	5.5	1827	8	ADN22715	Adn22715 Bacterial
44	123	5.5	454	5	AAU91510	Aau91510 Outer sur
45	123	5.5	1017	7	ADJ68352	Adj68352 Human hea

ALIGNMENTS

RESULT 1
AAE24371
ID AAE24371 standard; protein; 442 AA.
XX
AC AAE24371;
XX
DT 04-OCT-2002 (first entry)
XX
DE Human E93 programmed cell death modulating protein.

XX Human; cancer; programmed cell death modulating protein; adenocarcinoma;
cellular apoptosis; leukaemia; acquired immunodeficiency syndrome; AIDS;
neurodegenerative disease; Alzheimer's disease; retinitis pigmentosa;
Parkinson's disease; myelodysplastic syndrome; cerebellar degeneration;
aplastic anaemia; ischaemic injury; myocardial infarction; stroke;
reflexion injury; toxin-induced disease; genetic immunodeficiency;
vaccine; gene therapy; lymphoma; cytostatic; melanoma; neuroprotective;
myeloma; neutropenic; vasotropic; immunostimulant; cerebroprotective;
cardiant; E93 protein.

XX Homo sapiens.

XX Key Location/Qualifiers
PH Domain 353..405
/note= "Conserved domain"
FT FT WO200234882-A2.
PN PN 02-MAY-2002.

XX 29-OCT-2001; 2001WO-US048053.

XX 27-OCT-2000; 2000US-0243865P.

XX (UTMA-) UNIV MARYLAND BIOTECHNOLOGY INST.

XX Baehrecke EH;

XX WPI; 2002-479717/51.

XX Novel programmed cell death modulating proteins, useful for treating or
preventing disorders associated with abnormal cell proliferation and
apoptosis such as cancer, stroke, Parkinson's disease, myocardial
infarction.

XX Claim 1; Fig 4; 89pp; English.

XX The present invention relates to novel programmed cell death modulating

CC proteins and polynucleotides encoding such proteins. Sequences of the
CC invention are useful to screen potential cellular apoptosis inhibiting
CC compounds to determine their use as therapeutic agents for treatment of
CC diseases associated with increased programmed cell death. They are also
CC useful for treating or preventing disorders associated with decrease in
CC apoptosis. Programmed cell death modulating sequences are useful for
CC treating or preventing cancer e.g. adenocarcinoma, leukaemia, lymphoma,
CC melanoma, myeloma. Inhibition of the activity of the sequences of the
CC invention are useful for treating disorders associated with increase in
CC cell death or apoptosis such as acquired immunodeficiency syndrome
CC (AIDS), neurodegenerative diseases (e.g., Alzheimer's disease, retinitis
CC pigmentosa, Parkinson's disease and cerebellar degeneration), ischaemic
CC injuries (e.g., myocardial infarction, stroke, reperfusion injury),
CC myelodysplastic syndromes (e.g., aplastic anaemia), toxin-induced
CC diseases and other infectious or genetic immunodeficiencies. Sequences of
CC the invention are used as vaccines and in gene therapy. The present
CC sequence is human E93 programmed cell death modulating protein
XX
SQ Sequence 442 AA;
..Query Match 100.0%; Score 2250; DB 5; Length 442;
Best Local Similarity 100.0%; Pred. No. 9.9e-175;
Matches 442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MKKMQFALEYISKSGKTQENRNGSTGPSIVCKSIQMQAENSLQEOBGLDLTVNRM 60
Db 1 MKKMQFALEYISKSGKTQENRNGSTGPSIVCKSIQMQAENSLQEOBGLDLTVNRM 60
Qy 61 QEQNTQGGVLDLSTKTKTSIKSESSICDPSSNSVAGRLHNRNEDYVERSAEFADGILL 120
Db 61 QEQNTQGGVLDLSTKTKTSIKSESSICDPSSNSVAGRLHNRNEDYVERSAEFADGILL 120
Qy 121 SKALKDTQSGALDINKAGILYGPQKTLHLHLALPAGKFPASFKNKTROPHDSYKDSK 180
Db 121 SKALKDTQSGALDINKAGILYGPQKTLHLHLALPAGKFPASFKNKTROPHDSYKDSK 180
Qy 181 ETCVAVLOKVALWARAQARTKSKNLLETSEIKFPTASTYLLHQLTLQKMTVTFEKNES 240
Db 181 ETCVAVLOKVALWARAQARTKSKNLLETSEIKFPTASTYLLHQLTLQKMTVTFEKNES 240
Qy 241 LQYETSNPTVOLKIPQLRVSVSVSKSQPDGSLDVMYQVSKTSVLEGSALQKLNILPK 300
Db 241 LQYETSNPTVOLKIPQLRVSVSVSKSQPDGSLDVMYQVSKTSVLEGSALQKLNILPK 300
Qy 301 QNKIECGPVTHSSVDSYFLHGDLSPLCLNSKNGTVVDTGSENTEDGLDRKDSKQPKKRG 360
Db 301 QNKIECGPVTHSSVDSYFLHGDLSPLCLNSKNGTVVDTGSENTEDGLDRKDSKQPKKRG 360
Qy 361 RYROYDHEIMEEALAMVMSGKMSVSKAQGIYGVPHSTLEYKVKERSGTLTKTPPKKRLRP 420
Db 361 RYROYDHEIMEEALAMVMSGKMSVSKAQGIYGVPHSTLEYKVKERSGTLTKTPPKKRLRP 420
Qy 421 DTGLYNTDSTGSGCKNSKPV 442
Db 421 DTGLYNTDSTGSGCKNSKPV 442
RESULT 2
ABGI7942
ID ABGI7942 standard; protein; 630 AA.
XX
AC ABGI7942;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #17933.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.

XX PD 11-OCT-2001.
XX 30-MAR-2001; 2001WO-USO08631.
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX N-PSDB; AAS82129.
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
PS Claim 20; SEQ ID NO 48301; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activities in
XX polypeptide and polynucleotide sequences have application in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX amino acid sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 630 AA;
Query Match 85.0%; Score 1913.5; DB 4; Length 630;
Best Local Similarity 84.1%; Pred. No. 4.7e-147;
Matches 392; Conservative 11; Mismatches 30; Indels 33; Gaps 4;
Qy 10 IEVYSKSGTKQEN-----RNGSIGPSIVCKSIQMQAENSLQEE--- 48
Db 165 MELLSSQHDKKVENKIIQTRFKRQETLFAMRNSSDPMFRQSLQIKRELASIDENFTRK 224
Qy 49 ---QEGPLDLTVNR-----MOEQNTQOG---DGVLDLSTKTKTSIKSESSICDPSSNS 96
Db 225 KYTESKRKLQTNNEISSSDKGEFFHEQGNWDGVLDLSTKTKTSIKSESSICDPSSNS 284
Qy 97 VAGRLHNRNEDYVERSAEFADGILLSKALKDIQSGALDINKAGILYGPQKTLHLHLALP 156
Db 285 VAGRLHNRNEDYVERSAEFADGILLSKALKDIQSGALDINKAGILYGPQKTLHLHLALP 344
Qy 157 AGKPASFKNKTROPHDSYKDSKSKETCAVLQKVALWARAQARTKSKNLLETSEIKFP 216
Db 345 AGKPASFKNKTROPHDSYKDSKSKETCAVLQKVALWARAQARTKSKNLLETSEIKFP 404
Qy 217 TASTYLHQLTLQKMTVTFEKNESLQYETSNPTVOLKIPQLRVSVSVSKSQPDGSLDVM 276
Db 405 TASTYLHQLTLQKMTVTFEKNESLQYETSNPTVOLKIPQLRVSVSVSKSQPDGSLDVM 464
Qy 277 YQVSKTSVLEGSALQKLNILPKQNKIECGPVTHSSVDSYFLHGDLSPLCLNSKNGTV 336
Db 465 YQVSKTSVLEGSALQKLNILPKQNKIECGPVTHSSVDSYFLHGDLSPLCLNSKNGTV 524

Thu Feb 24 10:11:18 2005

us-10-016-768a-8.rag

PI Penn SG, Rank DR, Hanzel DK;
XX WPI; 2004-119264/12.
XX New human genome-derived single exon nucleic acid probes useful for human
XX gene expression analysis, for identifying or characterizing alternative
XX splicing events, for assessing genomic alterations or as tools for
XX surveying tissues.
XX Claim 45; SEQ ID NO 33895; 80pp; English.
XX The invention relates to a nucleic acid probe for measuring human gene
XX expression, comprising any of the 27,400 fully defined nucleotide
XX sequences in the specification, or their complements or fragments, and
XX encoding at least 8 amino acids of any of the 888 amino acid sequences
XX fully defined in the specification. The probe is a single exon probe that
XX hybridizes under high stringency conditions to a nucleic acid molecule
XX expressed in human cells or tissues. Also included are a spatially-
XX addressable set of single exon nucleic acid probes for measuring human
XX gene expression (comprising a plurality of single exon nucleic acid
XX probes cited above, where each of the plurality of probes is separately
XX and addressably isolatable or amplifiable from the plurality), a single
XX exon microarray for measuring human gene expression, a method of
XX measuring human gene expression, a vector comprising the single exon
XX probe cited above, an ORF-encoded peptide comprising at least 8
XX contiguous amino acids of any of the above-mentioned amino acid
XX sequences (optionally with conservative amino acid substitutions), an
XX isolated antibody that binds specifically to a peptide cited above,
XX a method of selling and/or licensing single exon probes or microarrays to
XX a customer desiring to measure gene expression, a method of providing
XX human gene expression data by subscription, and a computer-readable
XX storage medium which contains a database having a plurality of records
XX (each record including data on the expression of a single exon probe
XX cited above). The probe, methods and apparatus are useful in gene
XX expression analysis. The probes may be used as tools for surveying
XX tissues to detect the presence of expressed messages that contain their
XX specific exon, or in constructing genome-derived single exon microarrays.
XX In addition, the probes are used in identifying and characterizing
XX alternative splicing events, in detecting and characterizing gross
XX alterations in the genomic locus that includes their exon, in assessing
XX smaller genomic alterations, in priming the synthesis of nucleic acids,
XX or in expressing the ORF-encoded peptide. The present sequence is a human
XX single exon probe protein of the invention. Note: The sequence data for
XX this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from USPTO at
XX seqdata.uspto.gov/sequence.html?DocID=20030194704
XX
SQ Sequence 277 AA;
Query Match 62.7%; Score 1411; DB 8; Length 277;
Best Local Similarity 100.0%; Pred. No. 1.4e-106;
Matches 277; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 100 RLHRNREYVERSAEADGALLSKALDIOGALDINKAGILYGIPOKTLHLLEALPAGK 159
DB 1 RLHRNREYVERSAEADGALLSKALDIOGALDINKAGILYGIPOKTLHLLEALPAGK 60
QY 160 PASFNKTRPHDSYSYKDSKETCAVLQKVALWARAQARTKSKLNLETSEIKFTPTAS 219
DB 61 PASFNKTRPHDSYSYKDSKETCAVLQKVALWARAQARTKSKLNLETSEIKFTPTAS 120
QY 220 TYLHQLTLQKMWTOFKENESLOYETSNPTVOLKIPOLRVSSVSKSQPDGSLLDVNVQV 279
DB 121 TYLHQLTLQKMWTOFKENESLOYETSNPTVOLKIPOLRVSSVSKSQPDGSLLDVNVQV 180
QY 280 SKTSVLEGSALQKLNILPKONKIECSGPVTHSSVDSYFLHGLDLSPLCLNSKNGTVDGT 339
DB 181 SKTSVLEGSALQKLNILPKONKIECSGPVTHSSVDSYFLHGLDLSPLCLNSKNGTVDGT 240
QY 340 SENTEDGLDRKDSKQPRKGRGRYQYDHEIMEEAIAM 376
DB 241 SENTEDGLDRKDSKQPRKGRGRYQYDHEIMEEAIAM 277

RESULT 5
ADA54381
ID ADA54381 standard; protein; 213 AA.
XX
AC ADA54381;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human protein, SEQ ID 1949.
XX
KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease.
OS Homo sapiens.
XX EPI293569-A2.
XX
PD 19-MAR-2003.
XX
PF 21-MAR-2002; 2002EP-00006586.
XX
PR 14-SEP-2001; 2001JP-00328381.
PR 24-JAN-2002; 2002US-0350435P.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Matsuho Y;
XX
DR WPI; 2003-395539/38.
DR N-PSDB; ADA52742.
XX
PT New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 14; SEQ ID NO 1949; 205pp; English.
XX
CC The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 213 AA;
Query Match 48.4%; Score 1090; DB 6; Length 213;
Best Local Similarity 99.5%; Pred. No. 1.5e-80;
Matches 212; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 230 WYTFQFKENESLOYETSNPTVOLKIPOLRVSSVSKSQPDGSLLDVNVQVSTSVLEGS 289
DB 1 WYTFQFKENESLOYETSNPTVOLKIPOLRVSSVSKSQPDGSLLDVNVQVSTSVLEGS 60
QY 290 ALQKLNILPKONKIECSGPVTHSSVDSYFLHGLDLSPLCLNSKNGTVDGTSENTEDGLDR 349
DB 61 ALQKLNILPKONKIECSGPVTHSSVDSYFLHGLDLSPLCLNSKNGTVDGTSENTEDGLDR 120
QY 350 KDSKQPRKGRGRYQYDHEIMEEAIAMVWSGKMSVSKAQGIYGVPHSTILEYVKERSGTL 409
DB 121 KDSKQPRKGRGRYQYDHEIMEEAIAMVWSGKMSVSKAQGIYGVPHSTILEYVKERSGTL 180
QY 410 KTFPPKKKLRLPDTGLYNMTDSTGTCCKNSKXPV 442
DB 181 KTFPPKKKLRLPDTGLYNMTDSTGTCCKNSKXPV 213
RESULT 6
ADM04876

```

ID ADM04876 standard; protein; 311 AA.
XX
AC ADM04876;
XX
XX 20-MAY-2004 (first entry)
XX
DE Human protein of the invention SEQ ID NO:3561.
XX
KW human; gene therapy; diagnostic marker; pharmaceutical.
XX
OS Homo sapiens.
XX
PN EP1347046-A1.
XX
PD 24-SEP-2003.
XX
PF 12-APR-2002; 2002EP-00008400.
XX
PR 22-MAR-2002; 2002JP-00137785.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isoqai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-723558/69.
DR N-PSDB; ADM02433.
XX
XX New polynucleotides and polypeptides are useful in gene therapy, for
PT developing a diagnostic marker or medicines for regulating their
PT expression and activity, or as a target of gene therapy.
XX
PS Claim 1; SEQ ID NO 3561; 305pp; English.
XX
CC The invention relates to a novel human polynucleotide and the encoded
CC polypeptide. A polynucleotide of the invention may have a use in gene
CC therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
CC as a primer for synthesizing the polynucleotide or as a probe for
CC detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are
CC useful in gene therapy, for developing a diagnostic marker or medicines
CC for regulating their expression and activity, or as a target of gene
CC therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
CC are useful as pharmaceutical agents. The present sequence represents a
CC protein sequence of the invention.
XX
SQ Sequence 311 AA;
XX
XX Query Match 22.3%; Score 502; DB 7; Length 311;
XX Best Local Similarity 89.5%; Pred. No. 2.5e-32;
XX Matches 102; Conservative 3; Mismatches 9; Indels 0; Gaps 0;
XX
QY 1 MKKMIQPAIEYISKGTQNRNIGPSIVCKSIQNMNAENSLQEEQGLDITVNRM 60
Db
154 MKKMIQPAIEYISKGTQNRNIGPSIVCKSIQNMNAENSLQEEQGLDITVNRM 213
QY 61 QSONTQQQDGVLDLSTKTSIKSESSICDPSSENSAGRLNRNEDYVERSAE 114
Db 214 QSONTQQQDGVLDLSTKTSIKSESSICDPSSENSAGRLNRNEDYVERSAE 267
XX
RESULT 7
ADN35086
ID ADN35086 standard; protein; 433 AA.
XX
XX ADN35086;
AC
XX
XX 18-NOV-2004 (first entry)
XX
DE Nuclear receptor transcriptional corepressor.
XX
KW co-repressor; nuclear receptor; ligand-dependent transactivation;
KW steroid hormone signal.
XX

```

```

XX Homo sapiens.
XX WO2004029247-A1.
XX
XX 08-APR-2004.
XX
XX 25-SEP-2003; 2003WO-CA001477.
XX
XX 26-SEP-2002; 2002US-0413602P.
XX
XX (TRAN-) CENT TRANSLATIONAL RES IN CANCER.
XX
XX White J, Fernandes I;
XX
XX WPI; 2004-316124/29.
XX N-PSDB; ADN35085.
XX
XX New isolated co-repressor polypeptide that is operably interactable with
XX a nuclear receptor to actively repress transcription of DNA, useful for
XX inhibiting ligand-dependent transactivation in a cell.
XX
XX Claim 6; SEQ ID NO 2; 74pp; English.
XX
XX The present invention relates to an isolated co-repressor polypeptide
XX that is operably interactable with a nuclear receptor to actively repress
XX transcription of DNA. The co-repressor polypeptide is useful for
XX inhibiting ligand-dependent transactivation in a cell by one of a class I
XX and class II nuclear receptor, for repressing nuclear-receptor mediated
XX transcription in a cell, for modulating steroid hormone signaling in a
XX cell, and for regulating gene expression in a cell. The polypeptide is
XX also useful in an assay to select, for therapeutic purposes, compounds
XX that modulate transcription of gene expression associated with the
XX polypeptide. The present sequence represents the nuclear receptor
XX transcriptional corepressor.
XX
XX Query Match 22.2%; Score 499.5; DB 8; Length 433;
XX Best Local Similarity 33.6%; Pred. No. 6.3e-32;
XX Matches 158; Conservative 66; Mismatches 143; Indels 103; Gaps 20;
XX
QY 1 MKKMIQPAIEYISKGTQNRNIGPSIVCKSIQNMNAENSLQEE 48
Db 1 MORMIQQAAYTSKNSSTQDPNQSLPKASPVTTSTTAATTQNVLSKLLMAD 60
QY 49 QSEGLDLTVNRMQEONTQGGVLDLSTKTKT---SIKSESSICDPSSENSAGRLNRN 104
Db 61 QSPDLTLVRKQSQSEFSEQ-DGVLDLSTKSKPCAGSTLSHSPGCSSTQGNRGPSPQY 119
QY 105 REDYVERSAEAFADGLLSKALKD-----IQSGALDINKAGILYIPQKTLHLLEALPAG 158
Db 120 RPD-----GLRSGDGVPPRSQDGTREGFGHSTSLKVLPA-----RSLQISELLSRN 167
QY 159 K-----PASFKNTRDFHDSYKQSKETCAVLQKVALWARAQAE-RTEKSKLN--- 206
Db 168 QLSTAASLPGSLQN-----HQQH-----LILSREASWAKPHYENLSRMKFRNG 213
QY 207 -LLETSEIKFPTASTYHLQTLQKMTQFKENESLOVETSNPTVOLKIPOLRVSVSXKS 265
Db 214 ALSNISDLPFLAENSAFFMQLQ-----AKQGGKGVSHSSPVDLKLIPQVRGMDLSWE 266
QY 266 QPDGSGLLDVMYQVSKTSVL-----EGSALQKLKMLPKQNKIEC--SGPVTHSSVDSYF 319
Db 267 SRTGD-----QYSYSLVWGSQTESALSCLKRALLPKQSRKSMLDAGP-----DSNG 313
QY 320 LHGDLSPCLNSKNGTVGTSNTEDGLDRKQSKQPRKRGYRYDHDHIMEEAIAMVMS 379
Db 314 SDAE-----QSTSGQPYPTSDQEGD-----PGSKQPRKRGYRYQVNSSEILBEAISVMS 363
QY 380 GKMVSQAQGIYGVPHSTLEYKVKERSGTLTTPPKKLLRL-----PDTCL 424
Db 364 GKMVSQAQGIYGVPHSTLEYKVKERSGTLTTPPKKLLRL-----PDTCL 424

```

RESULT 8					
ADE71284					
ID	ADE71284 standard; protein; 572 AA.				
XX					
XX	ADE71284;				
XX	29-JAN-2004 (first entry)				
XX	Novel human protein #38.				
DE	human; novel protein; drug.				
XX	Homo sapiens.				
XX	JF2002345493-A.				
PN	03-DEC-2002.				
XX	29-MAR-2001; 2002JP-00049046.				
PF	29-MAR-2001; 2001JP-00095524.				
PR	(KAZU-) ZH KAZUSA DNA KENYUSHO.				
XX	WPI; 2003-460885/44.				
XX	N-PSDB; ADE71222.				
DR					
XX	A gene and a protein encoded by it, used in drugs.				
PT					
XX	Disclosure; Page 176-178; 257pp; Japanese.				
PS	The invention comprises the amino acid and coding sequences of novel				
CC	human proteins. The DNA and protein sequences of the invention are used				
CC	in drugs. The present amino acid sequence represents a novel human				
CC	protein of the invention.				
XX					
SQ	Sequence 572 AA;				
	Query Match 22.2%; Score 499.5; DB 7; Length 572;				
	Best Local Similarity 33.6%; Pred. No. 9.4e-32;				
	Matches 158; Conservative 66; Mismatches 143; Indels 103; Gaps 20;				
QY	1 MKKMTRQFAEIVISKSQTQE-----NRNGS-----IGPSIVCKSIQMMAENSLOEE 48				
DB	140 MORMIQQFAAEYTSKNSSSTQDPSPNSTKNOSLPKASPVVTSPTAATTQNPLSKLLMAD 199				
QY	49 QSGPDLTLVRNMQEONTQCGVLDTLTKT----SIKESESSICDPSSENSVAGRLHRN 104				
DB	200 QDSPDLTVRKQSPESEQ-DGVLDLTKKSPCAGSTSLSHSPGCSSTQGNGRPGRPQSY 258				
QY	105 REDYVERSAEFADLLSKALD-----IQSGALDINKAGILYGIPQKTLLHLLEALPAG 158				
DB	259 RPD-----GLURSGDVPPRSLODGTREGFGHSLSLVPLA-----RSLQISELLSRN 306				
QY	159 K-----PAGFKNKTRDFHDSYSKYDSKETCAVLQKVALARAAQAE-RTEKSKLN--- 206				
DB	307 QLSTAASLGPSGLQN-----HGQH-----LILSREASWAKTHYEFNLISRMKFRNG 352				
QY	207 -LLETSEIKFPASTAYHOLTQKWTPQEKNESLOYETSNPTVOLKIPQLRVSVSVKS 265				
DB	353 ALSNI SDLPFLAENSAPPKVALQ-----AKODGKKOVSHSPVDULKIPQVRGMDLSWE 405				
QY	266 QPDGSGLLDWYQVSTTSYL-----EGSALQKLKNILPKONKITEC--SGPVTHSSVDSYVF 319				
DB	406 SRTGD-----QYYSSSLVMGSGTESALS KLRAILPKQRKSMULDGP-----DSWG 452				
QY	320 LHGDSLPLCLNSKNGTVDGTSENTEDGLDKDSQPKKRGRVROYDHEIMEEAIAVMVS 379				
DB	453 SDAE-----QSTSGQPYPTSDQEGD-----PGSKQPKKKGRVROYNSEILEEAISVMVS 502				
QY	380 GKMSVSKAQGIYGVPHSTLEYKYKERSGTLKTPPKKLRL-----PDTGLI 424				

Db 140 MQRMIOQFAEYTSKNSSTODPQNPSTKNQSLPKASPVTTPTAATQNPVLSKLLMAD 199
Qy 49 QEGPLDLTVNRMQEQNTQQGDGVLDLSTKXT-----SIKSEESICDPSSSENSVAGRLHRN 104
Db 200 QDSPDLTVRKSSQSEPEQ-DGVLDLSTKSPCAGSTSLSHSPCSTQCGRPSPSQY 258
Qy 105 REDYVERSAEFAGLLSKALKD-----IOSGALDINKAGILYGIPOKTLILLHLEALPAG 158
Db 259 RPD-----GLRSGDGVPRSLQDGTREGFGHSTSLKVPLA-----RSIQISEELSRN 306
Qy 159 K-----PASFNKTRDPHDSYSYKDSKETCAVLQKQWALWARAOAE-RTEKSKLN--- 206
Db 307 QLSAASLGRSGLQN-----HGQH-----LILSREASWAKPHYEFNLSRMKPRGNG 352
Qy 207 -LLETSEIKPPTASTYLHQTLMQWVQPKKKNESLQYETNSPTVOLKIPOLRVSSVSKS 265
Db 353 ALSNISDLPLAENSAPPKMAAQ-----AKQDGKDVSHSPVDLKIPOVRGMDLSWE 405
Qy 266 QPDGSGLLDVMYQVSKTSVY-----EGSALQKLKNTLPKQNKIEC--SGPVTHSSVDYSYF 319
Db 406 SRTGD-----QYSYSSLVMSQTESALSUKLRALPKOSRKSMLDAGP-----DSWG 452
Qy 320 LHGDLPLCLNSKNGTVDGTSENTDGLDRKDSQPKRGRVQYDHEIMEEAIAMVMS 379
Db 453 SDAE-----QSTSGQPYPTSDQEGD-----PGSKQPKRGRVQYNSSEILEAISVMS 502
Qy 380 GKMSVSKAQSIYGVPHSTLYKVKERSGTLTTPPKKRLR-----PDTGL 424
Db 503 GKMSVSKAQSIYGVPHSTLYKVKERLGTLPKPKKMKLMRSEGGPDVSV 552
RESULT 10
ID ABP32451
XX ABP32451 standard; protein; 104 AA.
AC ABP32451;
XX
DT 09-JUL-2002 (first entry)
XX
DE Human ORF1424 protein, SEQ ID NO:2848.
XX
KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
KW disease monitoring; cytokine; cell proliferation; cell differentiation;
KW immune modulation; haematopoiesis regulation; tissue growth;
KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
KW behaviour; cancer; proliferative disorder; neurological disorder;
KW cardiovascular disease; immune system disorder; organ transplantation;
KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;
KW vasotropic; antipruritic; antidiabetic; cytostatic; neutrophic;
KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
KW cardiant; hypotensive; antichyroid; antiinflammatory; immunomodulator;
KW dermatological; analgesic; virucide; antibacterial; fungicide.
XX
OS Homo sapiens.
XX
PN WO200190366-A2.
XX
PD 29-NOV-2001.
XX
PF 24-MAY-2001; 2001WO-US017076.
XX
PR 24-MAY-2000; 2000US-0206690P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Leach MD, Shinkets RA;
XX
DR WPI; 2002-106200/14.
DR N-PSDB; ABN76477.
XX
PT Novel human polypeptides and polynucleotides useful for diagnosing,

PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and disorders related to organ
XX transplantation.
PS Claim 10; Page 971-972; 250pp; English.
XX
CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins
CC designated ORF (Open reading frame) 1-4534, and sequences ABN75054-
CC ABN79587 represent cDNAs encoding them. The invention also encompasses
CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
CC polynucleotides, the recombinant production of ORFX proteins, antibodies
CC specific for ORFX proteins, methods for detecting ORFX polynucleotides and
CC polypeptides, methods of screening for modulators of ORFX expression or
CC activity, and methods of screening individuals for a predisposition to an
CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC range of biological activities, such as cytokine, cell proliferation,
CC cell differentiation, immune modulation, haematopoiesis regulation,
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic activity, thrombolytic activity,
CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
CC and antiinfective activity, and may also be involved in the determination
CC of bodily characteristics, fertility and behaviour. ORFX proteins,
CC nucleic acids and antibodies may be used in the treatment of cancers,
CC other proliferative disorders such as psoriasis and benign tumours,
CC neurological disorders such as epilepsy and Alzheimer's disease,
CC cardiovascular diseases, immune system disorders, disorders related to
CC organ transplantation, disorders of tissue growth and regeneration,
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
CC storage disease, and infectious diseases caused by viral, bacterial,
CC fungal and other pathogens. ORFX nucleic acids may also be used as a
CC source of primers and probes, in the detection of ORFX genomic sequences
CC or transcripts, in the identification and cloning of homologous
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
CC nucleic acids may additionally be used to produce transgenic animals
CC which may be useful for studying the function and/or activity of ORFX
CC protein, and in drug screening. The ORFX proteins may also be used as
CC immunogens to generate specific antibodies, which are useful in the
CC diagnosis, treatment and monitoring of ORFX-associated diseases
XX
SQ Sequence 104 AA;
Query Match 13.1%; Score 294.5; DB 5; Length 104;
Best Local Similarity 74.4%; Pred. No. 4.3e-16;
Matches 58; Conservative 7; Mismatches 8; Indels 5; Gaps 1;
Qy 352 SKQPKRGRVQYDHEIMEEAIAMVMSGKMSVSKAQSIYGVPHSTLYKVKERSGTLT 411
Db 7 SKQPKRGRVQYNSXTEAISVMSGKMSVSKAQSIYGVPHSTLYKVKERSGTLT 66
Qy 412 PPKKKLRL-----PDTGL 424
Db 67 PPKKKLRLSEGGPDVSV 84
RESULT 11
AAE24592
ID AAE24592 standard; protein; 53 AA.
XX
AC AAE24592;
XX
DT 04-OCT-2002 (first entry)
XX
DE Human E93 programmed cell death modulating protein conserved domain.
XX
KW Human; cancer; programmed cell death modulating protein; adenocarcinoma;
KW cellular apoptosis; leukaemia; acquired immunodeficiency syndrome; AIDS;
KW neurodegenerative disease; Alzheimer's disease; retinitis pigmentosa;
KW Parkinson's disease; myelodysplastic syndrome; cerebellar degeneration;
KW aplastic anaemia; ischaemic injury; myocardial infarction; stroke;
KW reperfusion injury; toxin-induced disease; genetic immunodeficiency;
KW vaccine; gene therapy; lymphoma; cytostatic; melanoma; neuroprotective;

KW myeloma; neutropenic; vasotropic; immunostimulant; cerebroprotective;
 KW cardiant; E93 protein.
 XX Homo sapiens.

OS WO200234882-A2.

PN 02-MAY-2002.

PD 29-OCT-2001; 2001WO-US048053.

PF 27-OCT-2000; 2000US-0243865P.

XX (UYMA-) UNIV MARYLAND BIOTECHNOLOGY INST.

XX Baehrecke EH;

XX WPI; 2002-479717/51.

XX Novel programmed cell death modulating proteins, useful for treating or
 PT preventing disorders associated with abnormal cell proliferation and
 PT apoptosis such as cancer, stroke, Parkinson's disease, myocardial
 PT infarction.

XX Claim 1; Fig 1; 88pp; English.

XX The present invention relates to novel programmed cell death modulating
 CC proteins and polynucleotides encoding such proteins. Sequences of the
 CC invention are useful to screen potential cellular apoptosis inhibiting
 CC compounds to determine their use as therapeutic agents for treatment of
 CC diseases associated with increased programmed cell death. They are also
 CC useful for treating or preventing disorders associated with decrease in
 CC apoptosis. Programmed cell death modulating sequences are useful for
 CC treating or preventing cancer e.g. adenocarcinoma, leukaemia, lymphoma,
 CC melanoma, myeloma. Inhibition of the activity of the sequences of the
 CC invention are useful for treating disorders associated with increase in
 CC cell death or apoptosis such as acquired immunodeficiency syndrome
 CC (AIDS), neurodegenerative diseases (e.g., Alzheimer's disease, retinitis
 CC pigmentosa, Parkinson's disease and cerebellar degeneration), ischaemic
 CC injuries (e.g., myocardial infarction, stroke, reperfusion injury),
 CC myelodysplastic syndromes (e.g., aplastic anaemia), toxin-induced
 CC diseases and other infectious or genetic immunodeficiencies. Sequences of
 CC the invention are used as vaccines and in gene therapy. The present
 CC sequence is human E93 programmed cell death modulating protein conserved
 CC domain

XX Sequence 53 AA;

Query Match 12.1%; Score 273; DB 5; Length 53;
 Best Local Similarity 100.0%; Pred. No. 9.4e-15;
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 353 KQPRKRGYQYDHEIMEEIAIMVMSGKMSVSKAQIYGVPHSTLEYKVKER 405
 DB 1 KQPRKRGYQYDHEIMEEIAIMVMSGKMSVSKAQIYGVPHSTLEYKVKER 53

RESULT 12

AAE24593

ID AAE24593 standard; protein; 54 AA.

XX AAE24593;

XX 04-OCT-2002 (first entry)

XX Fish E93 programmed cell death modulating protein conserved domain.

XX Fish; cancer; programmed cell death modulating protein; adenocarcinoma;
 KW cellular apoptosis; leukaemia; acquired immunodeficiency syndrome; AIDS;
 KW neurodegenerative disease; Alzheimer's disease; retinitis pigmentosa;
 KW Parkinson's disease; myelodysplastic syndrome; cerebellar degeneration;
 KW aplastic anaemia; ischaemic injury; myocardial infarction; stroke;
 KW reperfusion injury; toxin-induced disease; genetic immunodeficiency;

KW vaccine; gene therapy; lymphoma; cytostatic; melanoma; neuroprotective;
 KW myeloma; neutropenic; vasotropic; immunostimulant; cerebroprotective;
 KW cardiant; E93 protein.

OS Tetraodon nigroviridis.

XX WO200234882-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-US048053.

XX 27-OCT-2000; 2000US-0243865P.

XX (UYMA-) UNIV MARYLAND BIOTECHNOLOGY INST.

XX Baehrecke EH;

XX WPI; 2002-479717/51.

XX Novel programmed cell death modulating proteins, useful for treating or
 PT preventing disorders associated with abnormal cell proliferation and
 PT apoptosis such as cancer, stroke, Parkinson's disease, myocardial
 PT infarction.

XX Claim 1; Fig 1; 88pp; English.

XX The present invention relates to novel programmed cell death modulating
 CC proteins and polynucleotides encoding such proteins. Sequences of the
 CC invention are useful to screen potential cellular apoptosis inhibiting
 CC compounds to determine their use as therapeutic agents for treatment of
 CC diseases associated with increased programmed cell death. They are also
 CC useful for treating or preventing disorders associated with decrease in
 CC apoptosis. Programmed cell death modulating sequences are useful for
 CC treating or preventing cancer e.g. adenocarcinoma, leukaemia, lymphoma,
 CC melanoma, myeloma. Inhibition of the activity of the sequences of the
 CC invention are useful for treating disorders associated with increase in
 CC cell death or apoptosis such as acquired immunodeficiency syndrome
 CC (AIDS), neurodegenerative diseases (e.g., Alzheimer's disease, retinitis
 CC pigmentosa, Parkinson's disease and cerebellar degeneration), ischaemic
 CC injuries (e.g., myocardial infarction, stroke, reperfusion injury),
 CC myelodysplastic syndromes (e.g., aplastic anaemia), toxin-induced
 CC diseases and other infectious or genetic immunodeficiencies. Sequences of
 CC the invention are used as vaccines and in gene therapy. The present
 CC sequence is fish E93 programmed cell death modulating protein conserved
 CC domain

XX Sequence 54 AA;

Query Match 10.4%; Score 233.5; DB 5; Length 54;
 Best Local Similarity 81.5%; Pred. No. 1.6e-11;
 Matches 44; Conservative 7; Mismatches 2; Indels 1; Gaps 1;

QY 353 KQPRKRGYQYDHEIMEEIAIMVMSGKMSVSKAQIYGVPHSTLEYKVKER 405
 DB 1 KQPRKRGYQYDHEIMEEIAIMVMSGKMSVSKAQIYGVPHSTLEYKVKER 54

RESULT 13

AAE24594

ID AAE24594 standard; protein; 53 AA.

XX AAE24594;

XX 04-OCT-2002 (first entry)

XX Mouse E93 programmed cell death modulating protein conserved domain.

XX Mouse; cancer; programmed cell death modulating protein; adenocarcinoma;
 KW cellular apoptosis; leukaemia; acquired immunodeficiency syndrome; AIDS;
 KW neurodegenerative disease; Alzheimer's disease; retinitis pigmentosa;
 KW Parkinson's disease; myelodysplastic syndrome; cerebellar degeneration;
 KW aplastic anaemia; ischaemic injury; myocardial infarction; stroke;

KW reperfusion injury; toxin-induced disease; genetic immunodeficiency;
 KW vaccine; gene therapy; lymphoma; cytostatic; melanoma; neuroprotective;
 KW myeloma; neurotropic; vasotropic; immunostimulant; cerebroprotective;
 KW cardiant; E93 protein.

OS Mus musculus.

PN WO200234882-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-US048053.

XX 27-OCT-2000; 2000US-0243865P.

XX (UYMA-) UNIV MARYLAND BIOTECHNOLOGY INST.

XX Baehrecke EH;

DR WPI; 2002-479717/51.

XX Novel programmed cell death modulating proteins, useful for treating or
 PT preventing disorders associated with abnormal cell proliferation and
 PT apoptosis such as cancer, stroke, Parkinson's disease, myocardial
 PT infarction.

PS Claim 1; Fig 1; 88pp; English.

XX The present invention relates to novel programmed cell death modulating
 CC proteins and polynucleotides encoding such proteins. Sequences of the
 CC invention are useful to screen potential cellular apoptosis inhibiting
 CC compounds to determine their use as therapeutic agents for treatment of
 CC diseases associated with increased programmed cell death. They are also
 CC useful for treating or preventing disorders associated with decrease in
 CC apoptosis. Programmed cell death modulating sequences are useful for
 CC treating or preventing cancer e.g. adenocarcinoma, leukaemia, lymphoma,
 CC melanoma, myeloma. Inhibition of the activity of the sequences of the
 CC invention are useful for treating disorders associated with increase in
 CC cell death or apoptosis such as acquired immunodeficiency syndrome
 CC (AIDS), neurodegenerative diseases (e.g., Alzheimer's disease, retinitis
 CC pigmentosa, Parkinson's disease and cerebellar degeneration), ischemic
 CC injuries (e.g., myocardial infarction, stroke, reperfusion injury),
 CC myelodysplastic syndromes (e.g., aplastic anaemia), toxin-induced
 CC diseases and other infectious or genetic immunodeficiencies. Sequences of
 CC the invention are used as vaccines and in gene therapy. The present
 CC sequence is mouse E93 programmed cell death modulating protein conserved
 CC domain

XX Sequence 53 AA;

Query Match 10.2%; Score 229; DB 5; Length 53;

Best Local Similarity 81.1%; Pred. No. 3.6e-11;

Matches 43; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 353 KQPRKRGYQYDHEIMEEALAMVSGKMSVSKAAGIYGVPHSTLEYKVKER 405

DB 1 KHPRKRGYQYNSILEPISVLMGKMSVSKAAGIYGVPHSTLEYKVKER 53

RESULT 14

ABB71145

ID ABB71145 standard; protein; 1140 AA.

XX ABB71145;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 40227.

XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL15248.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.

XX Disclosure; SEQ ID NO 40227; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB2072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1140 AA;

Query Match 8.9%; Score 200.5; DB 4; Length 1140;

Best Local Similarity 21.7%; Pred. No. 6.2e-07;

Matches 92; Conservative 66; Mismatches 146; Indels 119; Gaps 14;

QY 19 TCENRNGSIGPSIVCKSIQMNQAEISLQEQSGPLDITVNRMOEQNTQGGDGLDL--ST 76

DB 497 SOENENGNASLLQQQHQHQQHQQHQQHQQHQQHQQHQQHQQHQQHQQHQQHQQH 556

QY 77 KITSIKSEESSICDPSENSVAGRLHRNREDYVERSAEFADGLLSKALDKIQSGALDINK 136

DB 557 EDPILKIPSKVSGPSSSS-----LSP 579

QY 137 AGILYIPIOKTLLHLEALPAGKPASFKNTRDFHDSYK-----SKETCAVLQKVALW 192

DB 580 GGLVGG-----HHHPLNNNSLSISNNSN--HSSNSHRNGSNRSPHSASPLAAAV-- 628

QY 193 ARAQAERTEKSKLNLETSEIKFPTASTYHLQTLQKMYTQFKEKNESLQYETSNFTVOL 252

DB 629 --AQQGYSGNSLLTSSSSSIQMMASNIQRQI-----NEQSGQES----- 667

QY 253 KIPQLRVSSVSKSQPDGSGLLDMYQ-----VSKTSSVLEGSALQKLNLPKQNKICS 307

DB 668 ----LRNGNVSDCSNNGSGSSSLGYKKPSISVAKIIGGTDTSRFGASPNLLSQH----- 718

QY 308 GPVTHSSVDSPFLHGDLSPLCLNSKNGTVGDTSENTEGDLDRKDS--KQPRKRGYRQY 365

DB 719 ----HS---AHHL-----THQQQQQQLSAQALGKGTGPRKGRYRNY 753

QY 366 DHEIMEEALAMVSGKMSVSKAAGIYGVPHSTLEYKVKERSGTLKTPPKKLRLLPTGLY 425

DB 754 DRDSLVEAVKAVQRGEMSVHRAGSYGVPHSTLEYKVKERH---LMRPRKREPQPDV 810

QY 426 NMT 428

DB 811 GLT 813

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 24, 2005, 01:45:39 ; Search time 44 Seconds
(without alignments)
749.884 Million cell updates/sec

Title: US-10-016-768A-8
Perfect score: 2250
Sequence: 1 MKWMRQFAIEYISGKTKQ.....GLYNTDSTGSGCKNSKRPV 442

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/prodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/prodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/prodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/prodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	131.5	5.8	2375	4	US-09-538-092-1131
2	130	5.8	3248	1	Sequence 1131, Ap
3	130	5.8	3248	5	Sequence 1, Appli
4	127	5.6	534	4	PCT-US95-16216-1
5	127	5.6	1753	4	Sequence 184, App
6	126.5	5.6	750	4	US-09-538-092-184
7	124.5	5.5	2954	4	Sequence 19154, A
8	123	5.5	2482	1	Sequence 3544, Ap
9	123	5.5	3210	4	Sequence 1, Appli
10	122	5.4	1164	4	Sequence 6, Appli
11	122	5.4	1938	4	Sequence 1154, Ap
12	122	5.4	1959	4	Sequence 399, App
13	122	5.4	2781	4	Sequence 1847, Ap
14	122	5.4	2907	4	Sequence 8134, Ap
15	119.5	5.3	475	4	Sequence 10, Appli
16	118	5.2	924	4	Sequence 1, Appli
17	118	5.2	1147	4	Sequence 14568, A
18	117	5.2	1146	3	Sequence 18798, A
19	117	5.2	1940	4	Sequence 1074, A
20	117	5.2	1963	4	Sequence 6, Appli
21	115.5	5.1	868	4	Sequence 901, App
22	114.5	5.1	569	4	Sequence 8886, Ap
23	114.5	5.1	1279	3	Sequence 9731, Ap
24	114.5	5.1	1279	4	Sequence 18, Appli
25	114.5	5.1	1279	4	Sequence 2, Appli
26	114.5	5.1	1786	4	Sequence 2, Appli
27	114.5	5.1	2476	4	Sequence 7880, Ap
					Sequence 7, Appli

28	114	5.1	1260	4	US-09-248-796A-20258	Sequence 20258, A
29	113.5	5.0	431	4	US-09-286-981B-3	Sequence 3, Appli
30	113.5	5.0	905	4	US-09-248-796A-16333	Sequence 16333, A
31	113	5.0	593	4	US-09-270-767-41549	Sequence 41549, A
32	113	5.0	876	4	US-09-764-176-10	Sequence 10, Appli
33	113	5.0	1531	4	US-09-949-016-6740	Sequence 6740, Ap
34	112.5	5.0	1444	4	US-09-949-016-9652	Sequence 9652, Ap
35	112.5	5.0	2004	4	US-09-538-092-1371	Sequence 1371, Ap
36	112.5	5.0	2004	4	US-09-949-016-6756	Sequence 6756, Ap
37	111.5	5.0	1564	4	US-10-144-198-2	Sequence 2, Appli
38	111.5	5.0	1564	4	US-10-144-198-4	Sequence 4, Appli
39	111.5	5.0	1780	4	US-09-949-016-6899	Sequence 6899, Ap
40	111.5	5.0	1972	4	US-08-875-435B-3	Sequence 3, Appli
41	111	4.9	1589	3	US-08-755-587-189	Sequence 189, App
42	111	4.9	1786	3	US-08-973-462-8	Sequence 8, Appli
43	111	4.9	2753	4	US-09-949-016-7659	Sequence 7659, Ap
44	111	4.9	2753	4	US-09-949-016-7660	Sequence 7660, Ap
45	111	4.9	3924	4	US-09-538-092-1246	Sequence 1246, Ap

ALIGNMENTS

RESULT 1
US-09-538-092-1131
; Sequence 1131, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurapatSeqFormat Version 0.9
; SEQ ID NO 1131
; LENGTH: 2375
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: Polypeptide Accession Number P46100
US-09-538-092-1131

Query Match	5.8%	Score 131.5;	DB 4;	Length 2375;
Best Local Similarity	23.0%;	Pred. No. 0.02;		
Matches	100;	Conservative 63;	Mismatches 173;	Indels 99; Gaps 21;
QY	14	SKSGKTQENRNG-STGPSIVCKSIQWQAENSLOEQEGPLDITVNRMEQNTQQSGVL	72	
Db	667	STSGSDFDTKKGSAKSIISKKRTQSESS---NYDSELEKIKSMKIGAAR-----	718	
QY	73	DLSTKK--TSIKSESSICDPSSNSVAGRLHNRREDYVERSAEPADGLLSKALKDIQSG	130	
Db	719	--TTKRIINTKDFDSEDEKHSKGMNQGHKNLKTSGEGSDDAERQERTFFSAEG	776	
QY	131	ALDINKAGILYGIPTKLLHLH-EALPAGKPFASFKNTRDFHDSYSYKDSKETCAVLQKV	189	
Db	777	TVD-----KDTTLMELRDLRPPKQAS---ASTDGVDKLSGKSQSFSLVRKV	822	
QY	190	ALWARAQARTSKKUNLLETSEIPPTASTYLHQLTQKMTVPQFKENESLOYETSNPT	249	
Db	823	-----AETKEKSK-----HLKTKTKKV--QDGLSDIAEKFLKQDS--DETSED	864	
QY	250	VQLKIPQLRVSSVSKSQPDGSLDVMYOVSKTSSVLEGSALQKLNILPKQNKIESGP	309	
Db	865	KX-----QSKKGTEERKKPS-----DFKKVKVIRMEQYESSSDGTEK--LPEREI-CHFP	912	

QY 310 VTHSSVDSYFLRGDLSPLCLNSKNGTVGDSNTEDGLDRKDSKQPRKGRYQYDHEI 369
DB 913 KGIKOI-----KNGTTDG-----EKSKKIRDTSKKDELSDY 946
QY 370 MEAIAMVSGKMSVSK--AQGIYGVPHSTLBYKVKERSGTLTKPPKKK--LRLPDTGLY 425
DB 947 ABKSTGKSDSCSSDKSKNGAYG-----REKRCCLKGSRKQDCSSSDTEKY 998
QY 426 NMTDSTGSGCKNSK 440
DB 999 SMKEDG---CNSSDK 1010
RESULT 2
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
US-08-353-700-1
Query Match 5.8%; Score 130; DB 1; Length 3248;
Best Local Similarity 20.2%; Pred. No. 0.045;
Matches 104; Conservative 74; Mismatches 200; Indels 138; Gaps 21;
QY 7 QFAIEYISGKTQNR---NGSIGPSIVCKSIQNNQAEISLQEBQEGPLDLTVNRMQEQ 63
DB 2676 QDTLEVIQSSYKNLENELELTQWDKMSFVEKVNKTAKETELQREHWEAQTAEQLQEL 2735
QY 64 NTQQGQGVLDLSTKTKTSIKSESSICDPSENS-VAGRLHNRREDYVERSAFADGL--- 119
DB 2736 SGEKNRFLAGLQLLLEIKSKDKQLKELTLENSLCKSLDCMKDQVEKGVREIEAY 2795
QY 120 ---LSKALKDIOSGALDINKAGILYGPQKTLILLHLEALPAGKPA5FKNKTDPHDSYSY 176

DB 2796 QLRHAEAKKHQALLDITNKQ---YVEVIQT-----YREKL----- 2828
QY 177 KDSKEICAVLOKVALWARAQABTERKSLN--LLETSEIKFPTASTYHLQHLTQKMTQF 234
DB 2829 -TSKECLSSQKLEI---DLLKSSKEELNNSLKATQIILEELKTKMDNL---KYVNQL 2880
QY 235 KEKNESL-----QYETSNPTVQKIPQLRVSSVSKSQPDGSLDDVMYQVSKT 282
DB 2881 KKENERAQGRMKLLIKSKQLEBEKEILQKELSQAQAQ-----EKQKT 2924
QY 283 SSVLE-----GSALQKLNILPKQNKIE-----CSGPVTHSSVD--SYFLHGDLSPL 327
DB 2925 GTVMDTKVDELTTTEIKELKETLEKTEADYLDKVCSSLISHEKLEKAKEMLETQVAHL 2984
QY 328 C-----LNSKNGTVDG-----TSENTEDGLDRKDSKQPRKK-----RGRYQ 364
DB 2985 CSQSKQDSRGSPLGPVPGSPIPSVTEKRLSSGQNKASKGRQSSGIWENGRTTPA 3044
QY 365 YDHEIMEEAIAMVSGKMSVSKAQGI---YGVPH-----STLEYKVKERSGT 408
DB 3045 TPESFSKSKKAVNGIHPAEDTEGTEFEPEGLPEVVYKGFADIPITGKTSPIILRTTMA 3104
QY 409 LKTPPK---KKRLPDTGLYNMVTSNGSGCKNSKP 441
DB 3105 TRTSPRLAAQKALSPLSL-----GKENLAESSKP 3134
RESULT 3
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Ratner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-16216-1

Query Match	5.8%;	Score 130;	DB 5;	Length 3248;
Best Local Similarity	20.24;	Pred. No. 0.045;		
Matches 104;	Conservative 74;	Mismatches 200;	Indels 138;	Gaps 21;

QY	7	QFAIEYISKGKTQENR---	NGSIGPSIVCKSIQMNQAENSLOEQEGPDLDTVNRMOEQ	63
DB	2676	QDTLEVLQSSYKNLENELELT	KMDKMSFVEKVNMTAKETELOREHMHMAQKTAELQEEL	2735
QY	64	NTQOQGVLDLSTYKTSIKSEESI	CDPSSENS-VAGRLHRNREDYVRSAEFADGL---	119
DB	2736	SGEKNRLAGELQLLLEEIKSKDQ	KLKELTLENSSELKSKLDCMHKQDVEKEGKVREBIAEY	2795
QY	120	---LSKALDKDIQSGALDINKAGI	LYGPQKTLILLHLEALPAGKYPASFKNTRDFHDSYSY	176
DB	2796	QLRLHBAEKKHQALLDNTNQ---	YEVEIQT-----YREXL-----	2828
QY	177	KDSKETCAVLQKVALWARAQAE	RTSKLN--LLETSEIKFPTASTVYHOLTQLQKMTQF	234
DB	2829	-TSKEECLSSQKLEI-----	DLKSSKEELNNLSKATQIILEELKKTQMDNL---	2880
QY	235	KEKNESL-----	QYETSNPTVQUKIPQLRVSSVSKSQPDGSGLLDVMYQVSKT	282
DB	2881	KKENARAQGRMKLLIKSKCQLE	BEKEITLQKELSQLQAAQ-----EKQKT	2924
QY	283	SSVLE-----	GSALQKLNILPKQNKIE-----CSGPVTHSSVD--SYFLHGDLSPL	327
DB	2925	GTVMQTKVDBLTTEIKELTEK	TEADYEYLDKYCSLLISHKLEKAKEMLETOVAHL	2984
QY	328	C-----LNSKNGTVDG-----	TSENTEDGLDRKDSQPRKK-----RGRYRQ	364
DB	2985	CSQOSQDSRGSPLLGPVPGSP	IPISVTBKRLSSGQNKASGRQKRSSGIWNGRQPTPA	3044
QY	365	YDHEIMEEAIFAMVMSGMSVSK	AQGI-----YGVPH-----STLEYKVKERSGT	408
DB	3045	TPESFKSKSKKAVMSGIHPAED	TEGTEFEPEGLPEVVYKGFADIPTKTSPYILRRTTWA	3104
QY	409	LKTPPK---KKURLPDTGLNMTD	SGTGSCNKSXP 441	
DB	3105	TRTSPRLAAQKLALSPLSL---	-----GKENLAESSKP 3134	

RESULT 4

US-09-538-092-184

; Sequence 184, Application US/09538092

; Patent No. 6753314

; GENERAL INFORMATION:

; APPLICANT: Giot, Loic

; APPLICANT: Mansfield, Traci A.

; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same

; FILE REFERENCE: 15966-542

; CURRENT APPLICATION NUMBER: US/09/538,092

; CURRENT FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: 60/127,352

; PRIOR FILING DATE: 1999-04-01

; PRIOR APPLICATION NUMBER: 60/178,965

; PRIOR FILING DATE: 2000-02-01

; NUMBER OF SEQ ID NOS: 1387

; SOFTWARE: CuraPatSeqFormat Version 0.9

; SEQ ID NO 184

; LENGTH: 534

; TYPE: PRT

; ORGANISM: Saccharomyces cerevisiae

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (0)...(0)

; OTHER INFORMATION: Polypeptide Accession Number YDR299W

US-09-538-092-184

Query Match	5.6%;	Score 127;	DB 4;	Length 534;
Best Local Similarity	21.4%;	Pred. No. 0.0052;		
Matches 92;	Conservative 61;	Mismatches 157;	Indels 120;	Gaps 16;

QY	5	IROFAIEYISKGKTQENRNGSI-----	GPSIVCKSIQMNQAENSLOEQEGPDLDTV	57
----	---	-----------------------------	-------------------------------	----

QY 315 VDSYFLHGLDLSPLCLNSKNGTVDGTSNTDGLDRKD-----SKQPRKKRGYRQYDHE 368
Db 859 -----FLRSNLELKEKEDTNSWNYNQKKAASLFEKQLETKSNYKKMEAD 905
QY 369 IMBEAIAVMNMGKMSVSKAGIYGVPHS-----TLEYKVKERSGTLKTPPKKK 416
Db 906 LQKE-----LQSAFNEINYLNGLAGKVPDRLLSRVELEKVKVSFQKLEKALEEK 956

RESULT 8
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-328-254-6

Query Match 5.5%; Score 123; DB 1; Length 2482;
Best Local Similarity 20.0%; Pred. No. 0.13;
Matches 103; Conservative 74; Mismatches 201; Indels 138; Gaps 21;

QY 7 QFAIEYISKSGKTOENR---NGSIGPSIVCKSIQMNQAEKSLQEEQEGPLDLTVNRMQEQ 63
Db 1948 QDTLEVLQSSYKXNLENELELTQMDKMSFVEKYNKMTAKETELQREHMAQKTAELQEBL 2007
QY 64 NTQOQDGVLDLSTKTSIKSEESSICDPSSENS-VAGRLHRNREDYVERSAEADGL--- 119
Db 2008 SGEKNRLAGELQLLLEEIKSSQDLKELTLENSELKSLDCMHKQDQVEKEGKVREBIAEY 2067
QY 120 ---LSKALKDIOQSGALDINKAGILYGIPOKTLILLHLEALPAGKPAFKNTRDFHDSYSY 176
Db 2068 QRLHAEAKKHQALLDNTNQ---YEVEIQT-----YREKL----- 2100
QY 177 KDSKETCAVLQKVALWARAQAERTSKLN--LLETSEIKFPTASTYLTQLTKQMVTOF 234
Db 2101 -TSKEECLSSQKLEI-----DLKSSKEELNLSKATQILELKKTKMDNL---KYVNL 2152
QY 235 KEKNESL-----QYETSNPTVQLKIPQLRVSVSVKSPQDPGSLLDVYQVSKT 282

Db 2153 KKENERAQGMKLLIKSKQLEBEEKILOKELSOLQAAQ-----EKQKT 2196
QY 283 SSYLE-----GSALQKLNILPKQNKIE-----CSGPVTHSSVD--SYFLHGLDLSPL 327
Db 2197 GTVMDTKVDELTEIKELKETLEEKTEADEYLDKYCSLLISHEKLEKAKEMLETQVAHL 2256
QY 328 C-----LNSKNGTVDG-----TSNTEDEGLDRKDSKQPRKK-----RGYRQ 364
Db 2257 CSQSKQDSRGSPLLGPFVVPSPISPVTEKRLSSQNKASGRQRSSGIWGGGPTPA 2316
QY 365 YDHEIMEEAIAVMNMGKMSVSKAAGI-----YGVPH-----STLEYKVKERSGT 408
Db 2317 TPESFSKSKKXAMVSGIHPAEDTEGTEFEPGLPEVVKGFADIPGKTSPYLIRRTMA 2376
QY 409 LKTPPK---KKLRLPDTGLYNTMTDSTGTSCKKSSKP 441
Db 2377 TRTSPLAAQKLALSLSL-----GKENLAESSKP 2406

RESULT 9
US-09-538-092-1154
; Sequence 1154, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqformatter Version 0.9
; SEQ ID NO 1154
; LENGTH: 3210
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P49454
US-09-538-092-1154

Query Match 5.5%; Score 123; DB 4; Length 3210;
Best Local Similarity 20.0%; Pred. No. 0.2;
Matches 103; Conservative 74; Mismatches 201; Indels 138; Gaps 21;

QY 7 QFAIEYISKSGKTOENR---NGSIGPSIVCKSIQMNQAEKSLQEEQEGPLDLTVNRMQEQ 63
Db 2676 QDTLEVLQSSYKXNLENELELTQMDKMSFVEKYNKMTAKETELQREHMAQKTAELQEBL 2735
QY 64 NTQOQDGVLDLSTKTSIKSEESSICDPSSENS-VAGRLHRNREDYVERSAEADGL--- 119
Db 2736 SGEKNRLAGELQLLLEEIKSSQDLKELTLENSELKSLDCMHKQDQVEKEGKVREBIAEY 2795
QY 120 ---LSKALKDIOQSGALDINKAGILYGIPOKTLILLHLEALPAGKPAFKNTRDFHDSYSY 176
Db 2796 QRLHAEAKKHQALLDNTNQ---YEVEIQT-----YREKL----- 2828
QY 177 KDSKETCAVLQKVALWARAQAERTSKLN--LLETSEIKFPTASTYLTQLTKQMVTOF 234
Db 2829 -TSKEECLSSQKLEI-----DLKSSKEELNLSKATQILELKKTKMDNL---KYVNL 2880
QY 235 KEKNESL-----QYETSNPTVQLKIPQLRVSVSVKSPQDPGSLLDVYQVSKT 282
Db 2881 KKENERAQGMKLLIKSKQLEBEEKILOKELSOLQAAQ-----EKQKT 2924
QY 283 SSYLE-----GSALQKLNILPKQNKIE-----CSGPVTHSSVD--SYFLHGLDLSPL 327

Db 2925 GTVMDTKVDLTTTEIKELKELTLEETKEADEYLDKVCSSLISHEKLEKAKEMLETQVAHL 2984
QY 328 C-----LNSKNGTVDC-----TSENTEDGLDRKDSKQPRK-----RGRYRQ 364
Db 2985 CSQSQKDSRSGPLGPPVPGSPSPVTEKRLSSQNKASGKQKRSRSGIENGCGPTPA 3044
QY 365 YDHEIMEEAIAMVMSGKMSVSKAQGI-----YGVPH-----STLEYKVKERSGT 408
Db 3045 TPESFSKSGKAVMSGIHFAEDTGTETFEPEGLPEVVKKGADIPGTGKTSPIRLRTTMA 3104
QY 409 LKTPPK---KKLRLPTGLYNNMTDSTGSCCKSSKP 441
Db 3105 TRTSPRLAAQKALSPLSL-----GKNLAESSKP 3134

RESULT 10
US-09-538-092-399
; Sequence 399, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratPatSeqformat Version 0.9
; SEQ ID NO 399
; LENGTH: 1164
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YHR158C
US-09-538-092-399

Query Match 5.4%; Score 122; DB 4; Length 1164;
Best Local Similarity 19.7%; Pred. No. 0.05;
Matches 85; Conservative 69; Mismatches 159; Indels 118; Gaps 17;

QY 7 QFAIEYISKSGKTQENRNGSIGPSIVCKSIQMQAENSLQEQSGPLDLTVNRMQEONTQ 66
Db 694 QFKIKHYNESSELQN-----NTEIDKLS-----PVDITIKKSDTAGHD 733

QY 67 QGDGVLDSLTKK-----TSIKSESSICDPSSSENSVAGRLHRNREDYVERSABFAD 117
Db 734 SANHVIDASDEKYNVPMGDVPTDTKNEASV--PINRDATT-----EVVDEA----- 778

QY 118 GLLSKALDKIQSGALDINKAGILYGIPOKTLHLLEALPAGKPAFPKNTKTRDFHDSYSYK 177
Db 779 -LFEKLRSELOS-----LXELTHEKALEAG--AHIKELETWOLKLSQK 819

QY 178 DSKETCAV--LQKVALWARAQARTKSKLNLLTSEIKFPFTASTYLHQLTQKMTQFKREK 237
Db 820 NSGTKEIDELDSVRL-----QSKCEILEADNHSLEDKYNVELEFVNSKFLDIEN 869

QY 231 ---VTQFK--EKNESIQVETSPTVOLKIPOLRVSSVSKSQPDGGLLDVMYQVSKTSV 286
Db 870 LNEVIQFONEKIKLSLELE---PNYKELEELQIEHNLRSN----- 908

QY 287 EGSALQKLNILPKQNKIECSGPVTHS-----SVDSYFLHGLDPLCLNSKNGTVDGTSN 342
Db 909 -----ERLKNESKQHNEDIINNANVYSSQLGSLISHWKENRANGSFLESSSLISVDEN 963

QY 343 TEDGLDRKDSKQPRKR---GRYQYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLE 399
Db 964 GEKTVGEPYGDQSRHRRVINKLTNRLLDQLSRSQELTIS--KEKLSSEYHALKMEHSLS 1022

QY 400 YKVKERSGTLK 410
Db 1023 QDVLVKEKEIK 1033

RESULT 11
US-09-949-016-6417
; Sequence 6417, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6417
; LENGTH: 1938
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6417

Query Match 5.4%; Score 122; DB 4; Length 1938;
Best Local Similarity 17.3%; Pred. No. 0.11;
Matches 80; Conservative 87; Mismatches 185; Indels 110; Gaps 16;

QY 3 KMIROFAIEYISKSGKTQENRNGSIGPSIVCKSIQMQAENSLQEQSGPLDLTVNRMQE 62
Db 916 KILLEAKVKELTERLEEEEMNS-----ELVAKGNLEDKCSLKKDID-DLELTITKVEK 970

QY 63 QNTQQGDGVLDSLTKTKTSIKSESSICDPSSSENSVAGRLHRNREDYVERSABFADGLLSK 122
Db 971 EKHATENKYNLSEWTAL---EENISKLTKEKSLQEAHQOTLDDLQVEEDKVNGLIKI 1027

QY 123 ALK-----DIQSGALDINKAGILYGIPOKTLHLLEALPAGKPAFPKNTKTRDFHDSYSYK 177
Db 1028 NAKLEQOTDLEGSLE-----QEKKLRADLERAKRLEGLKMSQESIMDLENEK 1077

QY 178 DSKETCAVQLQKVALWARAQARTKSKLNLLTSEIKFPFTASTYLHQLTQKMTQFKREK 237
Db 1078 QQIE-----EKKKKEFELSOLQARIDDEQVHSLQFOKKIKELQAR 1118

QY 238 NESLOVE-TSNPTVOLKIPOLR-----VSSVSKSQPDGGLLDVMYQVSKTSV----- 285
Db 1119 IEELBEEIEAEHTLRAKIEKQSDLALELEIEISERLEEASGATSAQIEMNKKREAFQOM 1178

QY 286 ---LEGASLO-----KLKNILPKQNKIECSGPVTHSVDYSYFLH 321
Db 1179 RRDLEATLQHEATAATLKKQADSVAELGEQIDNLRVQKLEKEKSELKMEIDD--NA 1236

QY 322 GDLSPLCLNSKNGTVDGTSNTED---GLDRKDSKQPRKRGYRQYDHEI-MEEAIAMV 377
Db 1237 SNTLEALSCKSN--IERTCTVEDQSEIKAKDEQ-----TOLIHLDNNQKARLOT 1286

QY 378 MSGKMSVSKAQGIYGVPHSTLEYKVKERSGTLKTPPKKLR 419
Db 1287 QNGELS-----HRVEKEKSLISQLTKSKOAL 1312

RESULT 12
US-09-949-016-8134
; Sequence 8134, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:

```
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8134
; LENGTH: 1959
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-8134

Query Match      5.4%; Score 122; DB 4; Length 1959;
Best Local Similarity 17.3%; Pred. No. 0.11;
Matches 80; Conservative 87; Mismatches 185; Indels 110; Gaps 16;

QY 3 KMTIQFAIEYISGKTQENRNSIGPSIVCKSIQMNQAEISLOEBOEGPLDITVNRMQE 62
Db 937 KILLEAKVKELTLEERLEEEEMNS-----ELVAKRNLEDKCSLRDID-DLELTTKVEK 991
QY 63 QNTQOQDGVLDLSTKTSIKSESSICDPSSNSVAGRLHNRNEDYVERSAEFADGLLSK 122
Db 992 EKUATENKVNLESEMTAL---BENISKLTKEKKSQEAHQOITLDDLOVEEDKVGGLIKI 1048
QY 123 ALK-----DIQSGALDINKAGILYIPQKTLHLHLEALPAGKPAFKNKTRDFHDSYSYK 177
Db 1049 NAKLEQOTDDELSLE-----QEKLRADLERAKKLEGLDKWMSQESIMDLENDK 1098
QY 178 DSFETCAVLQKVALWARAQAEKSKLNLETSEIKFPTASTYVLHOLTLOKQVTFQEK 237
Db 1099 QQIE-----EKLKKEFELSQAQKIDDDQVHSLQFQKKIKELQAR 1139
QY 238 NESLOYE-TSNPTVOLKIPQLR-----VSSVSKSQPDGSGLLDVMYQVSKTSV----- 285
Db 1140 IELEEIEAEHTLRKIEKQSDRLARELEEISERLEEASGATSAQIEMNKKEAEFQKM 1199
QY 286 ---LEGSALQ-----KLKNILPKQNKIECSGPVTHSSHVSVDYSYFLH 321
Db 1200 RRDLEEATLQHEATAATLRKKQADSVAEELCEQIDNLRQVKQKLEKESKELKWEIDD--WA 1257
QY 322 GDLSPLCLNSKNGTVDGTSNTED---GLDRKDSKQPKRGYRQVDHEI-MEEAIAV 377
Db 1258 SNIEALSCKSN--IERTCTVDEQFSEIKAKDEQ-----TQLIHLNMQKARLQT 1307
QY 378 MSCKMSVSKAQGIYGVPHSTLEYKVKERSGTLTTPPKKLR 419
Db 1308 QNGELS-----HRVEEKESLISQLTQSKQAL 1333

RESULT 13
US-09-698-295-10
; Sequence 10, Application US/09698295
; Patent No. 6689584
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY FACTOR
; FILE REFERENCE: 06501-068001
; CURRENT APPLICATION NUMBER: US/09/698,295
; CURRENT FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: PCT/JP99/02340
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: JAPAN 10/137631
; PRIOR FILING DATE: 1998-04-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2907
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-698-295-1

Query Match      5.4%; Score 122; DB 4; Length 2907;
Best Local Similarity 19.7%; Pred. No. 0.21;
Matches 88; Conservative 86; Mismatches 180; Indels 92; Gaps 18;

QY 54 DLTVMRMEQNTQOQDGVLDLSTKK-----TSIKSESSICDPSSSE---NSVAGR 100
Db 846 EMDISKITEKKQDVVELLDSDSDKCKEPEVDDDMKTESHVNCQESSQVDVNVVSEG 905
QY 101 LHRNREDYVERSAEFADGLLSKALKDIQSGALDINKAGILYIPQKTL--LHLEALPAG 158
Db 906 FHLRTSVYKKTKSSKLDGLLERRIKQ-----FTLEEKQRLKIKLEGGIKG 951
QY 159 KPASFKNKTRDFHDSYSYKDSKETC---AVLQKVALWARAQAEKTE-----KSKLNL 208
Db 952 IGTSTNNSKNLSESPVITKAKEGCQSDSNRQES--PNANNDOPEDLIQGCQSQSSVL 1009
QY 209 ETSEIKFPTASTY-----LHOLTLOKQVTFQEKNESLOVETSNPTVOLKIPOL--RVSS 261
Db 1010 RMSDPHTTNKLYPKDRVLDVDSIRSPTKCPQN-SIEND-----IEEKVSDLASRQGE 1063
QY 262 VSKSQPDGSGLLDVMYQVSKTSVLE-GSALQKLNILPKQNK---IECSGPVTHSSHVS 317
Db 1064 PTKSKTKGN--DFFIDDSKLASADDITGLCKNKKPLIQEESDTIVSSSKSALHSSVPK 1120
QY 318 YFLHGLDPL-----CLNSKNGTVDGTS-----ENTEDGLDRKDSKQPKRG 360
Db 1121 STNDRDATPLSRAMDFEGLKCDSESNLENSSDTVSIQDSSEDMIVQNSNESISEQF 1180
QY 361 RVROYDHEIMEEAIAMVMSGMSVSKAQGIYGVPHSTLEYKVKERSGTLTTPPKKLR 420
Db 1181 RTRQDVEVLEPLKCELSVGEST-----GNCEBRLPVKGTENGKPKFSQKKLEERP 1232
QY 421 -----DTGLYNNMDSGTGSCKNSSK 440
Db 1233 VNKCSDQIKLKNITTDKKNENRESEK 1258

RESULT 14
US-09-698-295-1
; Sequence 1, Application US/09698295
; Patent No. 6689584
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY FACTOR
; FILE REFERENCE: 06501-068001
; CURRENT APPLICATION NUMBER: US/09/698,295
; CURRENT FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: PCT/JP99/02340
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: JAPAN 10/137631
; PRIOR FILING DATE: 1998-04-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2907
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-698-295-1

Query Match      5.4%; Score 122; DB 4; Length 2907;
Best Local Similarity 19.7%; Pred. No. 0.21;
Matches 88; Conservative 86; Mismatches 180; Indels 92; Gaps 18;

QY 54 DLTVMRMEQNTQOQDGVLDLSTKK-----TSIKSESSICDPSSSE---NSVAGR 100
Db 972 EMDISKITEKKQDVVELLDSDSDKCKEPEVDDDMKTESHVNCQESSQVDVNVVSEG 1031
```

```
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8134
; LENGTH: 1959
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-8134

Query Match      5.4%; Score 122; DB 4; Length 1959;
Best Local Similarity 17.3%; Pred. No. 0.11;
Matches 80; Conservative 87; Mismatches 185; Indels 110; Gaps 16;

QY 3 KMTIQFAIEYISGKTQENRNSIGPSIVCKSIQMNQAEISLOEBOEGPLDITVNRMQE 62
Db 937 KILLEAKVKELTLEERLEEEEMNS-----ELVAKRNLEDKCSLRDID-DLELTTKVEK 991
QY 63 QNTQOQDGVLDLSTKTSIKSESSICDPSSNSVAGRLHNRNEDYVERSAEFADGLLSK 122
Db 992 EKUATENKVNLESEMTAL---BENISKLTKEKKSQEAHQOITLDDLOVEEDKVGGLIKI 1048
QY 123 ALK-----DIQSGALDINKAGILYIPQKTLHLHLEALPAGKPAFKNKTRDFHDSYSYK 177
Db 1049 NAKLEQOTDDELSLE-----QEKLRADLERAKKLEGLDKWMSQESIMDLENDK 1098
QY 178 DSFETCAVLQKVALWARAQAEKSKLNLETSEIKFPTASTYVLHOLTLOKQVTFQEK 237
Db 1099 QQIE-----EKLKKEFELSQAQKIDDDQVHSLQFQKKIKELQAR 1139
QY 238 NESLOYE-TSNPTVOLKIPQLR-----VSSVSKSQPDGSGLLDVMYQVSKTSV----- 285
Db 1140 IELEEIEAEHTLRKIEKQSDRLARELEEISERLEEASGATSAQIEMNKKEAEFQKM 1199
QY 286 ---LEGSALQ-----KLKNILPKQNKIECSGPVTHSSHVSVDYSYFLH 321
Db 1200 RRDLEEATLQHEATAATLRKKQADSVAEELCEQIDNLRQVKQKLEKESKELKWEIDD--WA 1257
QY 322 GDLSPLCLNSKNGTVDGTSNTED---GLDRKDSKQPKRGYRQVDHEI-MEEAIAV 377
Db 1258 SNIEALSCKSN--IERTCTVDEQFSEIKAKDEQ-----TQLIHLNMQKARLQT 1307
QY 378 MSCKMSVSKAQGIYGVPHSTLEYKVKERSGTLTTPPKKLR 419
Db 1308 QNGELS-----HRVEEKESLISQLTQSKQAL 1333

RESULT 13
US-09-698-295-10
; Sequence 10, Application US/09698295
; Patent No. 6689584
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY FACTOR
; FILE REFERENCE: 06501-068001
; CURRENT APPLICATION NUMBER: US/09/698,295
; CURRENT FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: PCT/JP99/02340
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: JAPAN 10/137631
; PRIOR FILING DATE: 1998-04-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
```

```

Query Match      5.3%; Score 119.5; DB 4; Length 475;
Best Local Similarity 22.3%; Pred. No. 0.021;
Matches 96; Conservative 55; Mismatches 180; Indels 99; Gaps 16;

      y      39 NOAENS LQEEQGPDLDTW--RMQEQTQOQDGVLDLSTKTKTSIKBESSICDPSSENS 96
      :      :      :      :      :      :      :      :      :      :      :
      b      88 SQTQPS LQKRSAPVPGAVSSPNPTPSSQPSGERSUSQAPS--SSNDLLADPA---- 142

      y      97 VAGRLHRNREDYVERSAPFADGLLSKALKDIOSGALDINKAGILYGIQKTLHLALPALP 156
      :      :      :      :      :      :      :      :      :      :      :
      b      143 -----ISQLSQATSDI-----ANVSNQIKSLTSQTTLNHEKKTR 177

      y      157 AGK-----PASFNKTRDFHDYSYKDSKETCAVLQKVALWARAQAE-----RTEKSK 204
      :      :      :      :      :      :      :      :      :      :      :
      b      178 AEKELQRLITKSEIENIKQLRASV---DNEVIQVEQVEANLATAKEETEARSEAS-- 232

      y      205 LNLLETSEIKFPTASTYIHLQ-----TLQKMYTQFKENESLOYETSNPVTQLKIPQLRV 259
      :      :      :      :      :      :      :      :      :      :      :
      b      233 -----IAEAKVNSLGSELHEKVAMESLQENSTLKEKLGSLNAE-----IVELEK 278

      y      260 SSVYSKQPDGSGLLDVMYQVSKTSSVLEGSAQLKLNILPKQNKTECSGPVTHSVSDSYF 319
      :      :      :      :      :      :      :      :      :      :      :

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 24, 2005, 01:54:30 ; Search time 131 Seconds
(without alignments)
1104.126 Million cell updates/sec

Title: US-10-016-768A-8
Perfect score: 2250
Sequence: 1 MKKMTQFAIEYISKGTQ.....GLYNTDSTGCKNSKRPV 442

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1380268 seqs, 327241040 residues

Total number of hits satisfying chosen parameters: 1380268

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications AA.*
- 1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/prodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/prodata/1/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/prodata/1/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 13: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 14: /cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pep.*
 - 15: /cgn2_6/prodata/1/pubpaa/US10D_PUBCOMB.pep.*
 - 16: /cgn2_6/prodata/1/pubpaa/US10E_PUBCOMB.pep.*
 - 17: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*
 - 18: /cgn2_6/prodata/1/pubpaa/US11_NEW_PUB.pep.*
 - 19: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 20: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					Description	
Result No.	Score	Query Match	Length	ID		
1	2250	100.0	442	13	US-10-016-768-8	Sequence 8, Appli
2	1754	78.0	357	15	US-10-085-198-24	Sequence 24, Appl
3	1411	62.7	277	14	US-10-029-386-33895	Sequence 33895, A
4	1090	48.4	213	15	US-10-094-749-1949	Sequence 1949, Ap
5	502	22.3	311	15	US-10-108-260A-3561	Sequence 3561, Ap
6	499.5	22.2	572	16	US-10-408-785A-1766	Sequence 1766, Ap
7	294.5	13.1	103	11	US-09-864-408A-2848	Sequence 2848, Ap
8	273	12.1	54	13	US-10-016-768-2	Sequence 2, Appli
9	233.5	10.4	53	13	US-10-016-768-3	Sequence 3, Appli
10	229	10.2	53	13	US-10-016-768-4	Sequence 4, Appli
11	200.5	8.9	1165	13	US-10-016-768-10	Sequence 10, Appl
12	165	7.3	53	13	US-10-016-768-1	Sequence 1, Appli
13	163	7.2	53	13	US-10-016-768-5	Sequence 5, Appli

14	133.5	5.9	1109	16	US-10-387-767-8	Sequence 8, Appli
15	132.5	5.9	848	13	US-10-011-588-45	Sequence 45, Appl
16	132	5.9	787	16	US-10-437-963-125796	Sequence 125796,
17	131.5	5.8	870	14	US-10-029-386-32827	Sequence 32827, A
18	131.5	5.8	2375	16	US-10-408-765A-277	Sequence 277, App
19	131.5	5.8	2492	16	US-10-697-526-2	Sequence 2, Appli
20	131	5.8	1297	15	US-10-354-774-77	Sequence 77, Appl
21	131	5.8	1297	15	US-10-271-012-77	Sequence 77, Appl
22	131	5.8	1297	15	US-10-452-024-7	Sequence 7, Appli
23	131	5.8	1297	15	US-10-452-024-150	Sequence 150, App
24	131	5.8	1297	15	US-10-205-516-14	Sequence 14, Appl
25	131	5.8	1297	16	US-10-729-122-77	Sequence 77, Appl
26	131	5.8	1297	16	US-10-729-039-77	Sequence 77, Appl
27	131	5.8	2017	15	US-10-452-024-146	Sequence 146, App
28	127	5.6	534	10	US-09-893-519A-37	Sequence 37, Appl
29	126	5.6	433	15	US-10-369-493-2477	Sequence 2477, Ap
30	125.5	5.6	1819	15	US-10-335-977-7981	Sequence 7981, Ap
31	125.5	5.6	1820	15	US-10-335-977-7982	Sequence 7982, Ap
32	125	5.6	1307	15	US-10-205-516-28	Sequence 28, Appl
33	124.5	5.5	972	9	US-09-924-154-16	Sequence 16, Appl
34	123.5	5.5	1827	15	US-10-369-493-5368	Sequence 5368, Ap
35	123	5.5	1017	16	US-10-408-765A-158	Sequence 158, App
36	122.5	5.4	748	15	US-10-425-114-72422	Sequence 72422, A
37	122.5	5.4	958	15	US-10-369-493-6600	Sequence 6600, Ap
38	122.5	5.4	958	15	US-10-369-493-6601	Sequence 6601, Ap
39	122	5.4	454	15	US-10-369-339-70	Sequence 70, Appl
40	122	5.4	454	15	US-10-369-100-160	Sequence 160, App
41	122	5.4	1435	15	US-10-369-493-1390	Sequence 1390, Ap
42	122	5.4	1938	16	US-10-408-765A-1168	Sequence 1168, Ap
43	122	5.4	1938	16	US-10-408-765A-1629	Sequence 1629, Ap
44	122	5.4	2781	15	US-10-263-929-122	Sequence 122, App
45	121.5	5.4	1384	16	US-10-473-576-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1
US-10-016-768-8
; Sequence 8, Application US/10016768
; Publication No. US20020142443A1
; GENERAL INFORMATION:
; APPLICANT: Baehrcke, Eric H.
; TITLE OF INVENTION: GENES REGULATING PROGRAMMED CELL DEATH
; FILE REFERENCE: 4115-131
; CURRENT APPLICATION NUMBER: US/10/016,768
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 8
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-016-768-8

Query Match					100.0%;	Score	2250;	DB	13;	Length	442;
Best Local Similarity					100.0%;	Pred. No.	2e-169;				
Matches					442;	Conservative	0;	Mismatches	0;	Indels	0;
QY	1	MKKMTQFAIEYISKGTQENRNGSIGPSIVCKSIQMNQAE	NSLOE	QEGPLDLTVNRM	60						
DB	1	MKKMTQFAIEYISKGTQENRNGSIGPSIVCKSIQMNQAE	NSLOE	QEGPLDLTVNRM	60						
QY	61	QEQNTQQGVLGDLSTKTSIKSESSICDPSENSVAGLRHNR	EDYVERSA	EAFADGLL	120						
DB	61	QEQNTQQGVLGDLSTKTSIKSESSICDPSENSVAGLRHNR	EDYVERSA	EAFADGLL	120						
QY	121	SKALXDIQSGALDINKAGILYGPQKTLHLLEALPAGKPA	SPKNTKTR	PHDSYSYKDSK	180						
DB	121	SKALXDIQSGALDINKAGILYGPQKTLHLLEALPAGKPA	SPKNTKTR	PHDSYSYKDSK	180						
QY	181	ETCAVLQKVALWARQAERTESKLNLLTSEIKPTASTYLH	QLTQKMTQ	FKENES	240						
DB	181	ETCAVLQKVALWARQAERTESKLNLLTSEIKPTASTYLH	QLTQKMTQ	FKENES	240						

Db 181 ETCVALQKVALWARAQARTEKSKLNLETSEIKPTASTYHLQTLQKMTQFKEKNES 240
 QY 241 LOYETSNPTVLKIPOLRVSSVSKSOPDGSGLLDVWYQVSKTSSVLEGSALQKLNILPK 300
 Db 241 LOYETSNPTVLKIPOLRVSSVSKSOPDGSGLLDVWYQVSKTSSVLEGSALQKLNILPK 300
 QY 301 QNKIECSGPTVTHSSVDSYFLHGDLSPLCLNSKNGTVDTGSENTEGLDRKDSKQPRKRG 360
 Db 301 QNKIECSGPTVTHSSVDSYFLHGDLSPLCLNSKNGTVDTGSENTEGLDRKDSKQPRKRG 360
 QY 361 RYQYDHEIMEBAIAMVMSGKMSVSKAOGIYGVPHSTLEYKVKERSGTLTKTPPKKLRP 420
 Db 361 RYQYDHEIMEBAIAMVMSGKMSVSKAOGIYGVPHSTLEYKVKERSGTLTKTPPKKLRP 420
 QY 421 DTGLYNNMTSGTGCKNSKPKV 442
 Db 421 DTGLYNNMTSGTGCKNSKPKV 442

RESULT 2

US-10-085-198-24
 ; Sequence 24, Application US/10085198
 ; Publication No. US20040009907A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Alsobrook et al.
 ; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
 ; FILE REFERENCE: 21402-279
 ; CURRENT APPLICATION NUMBER: US/10/085,198
 ; CURRENT FILING DATE: 2002-02-25
 ; PRIOR APPLICATION NUMBER: 60/271,646
 ; PRIOR FILING DATE: 2001-02-26
 ; PRIOR APPLICATION NUMBER: 60/276,401
 ; PRIOR FILING DATE: 2001-03-16
 ; PRIOR APPLICATION NUMBER: 60/311,981
 ; PRIOR FILING DATE: 2001-08-13
 ; PRIOR APPLICATION NUMBER: 60/312,858
 ; PRIOR FILING DATE: 2001-08-16
 ; PRIOR APPLICATION NUMBER: 60/271,840
 ; PRIOR FILING DATE: 2001-02-27
 ; PRIOR APPLICATION NUMBER: 60/277,324
 ; PRIOR FILING DATE: 2001-03-20
 ; PRIOR APPLICATION NUMBER: 60/286,096
 ; PRIOR FILING DATE: 2001-04-21
 ; PRIOR APPLICATION NUMBER: 60/299,695
 ; PRIOR FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: 60/315,614
 ; PRIOR FILING DATE: 2001-08-29
 ; PRIOR APPLICATION NUMBER: 60/272,405
 ; PRIOR FILING DATE: 2001-02-28
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 653
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 24
 ; LENGTH: 356
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-085-198-24

Query Match 78.0%; Score 1754; DB 15; Length 356;
 Best Local Similarity 100.0%; Pred. No. 2.6e-130; Indels 0; Gaps 0;
 Matches 343; Conservative 0; Mismatches 0
 QY 100 RLHNRREDYVERSAEFADGLLSKALKDIQSGALDINKAGILYGIPOKTLHLLEALPAGK 159
 Db 14 RLHNRREDYVERSAEFADGLLSKALKDIQSGALDINKAGILYGIPOKTLHLLEALPAGK 73
 QY 160 PASFNKTRDFHDSYSYKDSKETCAVLQKVALWARAQARTEKSKLNLETSEIKFPTAS 219
 Db 74 PASFNKTRDFHDSYSYKDSKETCAVLQKVALWARAQARTEKSKLNLETSEIKFPTAS 133
 QY 220 TYLHQLTLQKMTQFKEKNESLOYETSNPTVLKIPOLRVSSVSKSOPDGSGLLDVWYQV 279
 Db 134 TYLHQLTLQKMTQFKEKNESLOYETSNPTVLKIPOLRVSSVSKSOPDGSGLLDVWYQV 193

QY 280 SKTSSVLEGSALQKLNILPKQNKIECSGPTVTHSSVDSYFLHGDLSPLCLNSKNGTVDTG 339
 Db 194 SKTSSVLEGSALQKLNILPKQNKIECSGPTVTHSSVDSYFLHGDLSPLCLNSKNGTVDTG 253
 QY 340 SENTEDGLDRKDSKQPRKRGRYQYDHEIMEBAIAMVMSGKMSVSKAOGIYGVPHSTLE 399
 Db 254 SENTEDGLDRKDSKQPRKRGRYQYDHEIMEBAIAMVMSGKMSVSKAOGIYGVPHSTLE 313
 QY 400 YKVKERSGTLTKTPPKKLRPDTGLYNNMTSGTGCKNSKPKV 442
 Db 314 YKVKERSGTLTKTPPKKLRPDTGLYNNMTSGTGCKNSKPKV 356

RESULT 3

US-10-029-386-33895
 ; Sequence 33895, Application US/10029386
 ; Publication No. US20030194704A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharron G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GI
 ; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
 ; FILE REFERENCE: AEONICA-X-2
 ; CURRENT APPLICATION NUMBER: US/10/029,386
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 34288
 ; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 33895
 ; LENGTH: 277
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AC005768.16
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.85
 ; OTHER INFORMATION: SWISSPROT HIT: Q9YID8, EVALUATE 1.60e+00
 US-10-029-386-33895

Query Match 62.7%; Score 1411; DB 14; Length 277;
 Best Local Similarity 100.0%; Pred. No. 2.5e-103; Indels 0; Gaps 0;
 Matches 277; Conservative 0; Mismatches 0
 QY 100 RLHNRREDYVERSAEFADGLLSKALKDIQSGALDINKAGILYGIPOKTLHLLEALPAGK 159
 Db 1 RLHNRREDYVERSAEFADGLLSKALKDIQSGALDINKAGILYGIPOKTLHLLEALPAGK 60
 QY 160 PASFNKTRDFHDSYSYKDSKETCAVLQKVALWARAQARTEKSKLNLETSEIKFPTAS 219
 Db 61 PASFNKTRDFHDSYSYKDSKETCAVLQKVALWARAQARTEKSKLNLETSEIKFPTAS 120
 QY 220 TYLHQLTLQKMTQFKEKNESLOYETSNPTVLKIPOLRVSSVSKSOPDGSGLLDVWYQV 279
 Db 121 TYLHQLTLQKMTQFKEKNESLOYETSNPTVLKIPOLRVSSVSKSOPDGSGLLDVWYQV 180
 QY 280 SKTSSVLEGSALQKLNILPKQNKIECSGPTVTHSSVDSYFLHGDLSPLCLNSKNGTVDTG 339
 Db 181 SKTSSVLEGSALQKLNILPKQNKIECSGPTVTHSSVDSYFLHGDLSPLCLNSKNGTVDTG 240
 QY 340 SENTEDGLDRKDSKQPRKRGRYQYDHEIMEBAIAM 376
 Db 241 SENTEDGLDRKDSKQPRKRGRYQYDHEIMEBAIAM 277

RESULT 4

US-10-094-749-1949
 ; Sequence 1949, Application US/10094749
 ; Publication No. US20030219741A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ISOGAI, TAKAO
 ; APPLICANT: SUGIYAMA, TOMOYASU
 ; APPLICANT: OTSUKI, TETSUJI
 ; APPLICANT: WAKAMATSU, AI

APPLICANT: SATO, HIROYUKI
APPLICANT: ISHII, SHIZUKO
APPLICANT: YAMAMOTO, JUN-ICHI
APPLICANT: ISONO, YUUKO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORU
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOHICO
APPLICANT: YOSHIKAWA, TSUTOMU
APPLICANT: OTSUKA, MOTYUKI
APPLICANT: NAGAHARI, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
FILE REFERENCE: 084335/0160
CURRENT APPLICATION NUMBER: US/10/094,749
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 60/350,435
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: JP 2001-328381
PRIOR FILING DATE: 2001-09-14
NUMBER OF SEQ ID NOS: 3381
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1949
LENGTH: 213
TYPE: PRT
ORGANISM: Homo sapiens
US-10-094-749-1949

Query Match 48.4%; Score 1090; DB 15; Length 213;
Best Local Similarity 99.5%; Pred. No. 4.2e-78;
Matches 212; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 230 MTQPKKNESLQYETSNPTVOLKIPQLRVSSVSKSQPDGSLDVMYQVSKTSSVLEGS 289
DB 1 MTQPKKNESLQYETSSFTVQKIPQLRVSSVSKSQPDGSLDVMYQVSKTSSVLEGS 60
QY 290 ALQKLKNILPKQNKIECSGPVTHSSVDSYFLHGDLSPLCLNSKNGTVDGTSNTDGLDR 349
DB 61 ALQKLKNILPKQNKIECSGPVTHSSVDSYFLHGDLSPLCLNSKNGTVDGTSNTDGLDR 120
QY 350 KDSQPKRKRGRYQYDHEIMEAIAVMGSKMSVSKAAGIYGVPHSTLEYKVKERSGTL 409
DB 121 KDSQPKRKRGRYQYDHEIMEAIAVMGSKMSVSKAAGIYGVPHSTLEYKVKERSGTL 180
QY 410 KTPPKKKLRLPDTGLNMTDSDGTCNCKSPV 442
DB 181 KTPPKKKLRLPDTGLNMTDSDGTCNCKSPV 213

RESULT 5
US-10-108-260A-3561
Sequence 3561, Application US/10108260A
Publication No. US20040005560A1
GENERAL INFORMATION:
APPLICANT: HELIX RESEARCH INSTITUTE
TITLE OF INVENTION: No. US20040005560A1el full length cdna
FILE REFERENCE: H1-A0106
CURRENT APPLICATION NUMBER: US/10/108,260A
CURRENT FILING DATE: 2002-03-27
NUMBER OF SEQ ID NOS: 5458
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3561
LENGTH: 311
TYPE: PRT
ORGANISM: Homo sapiens
US-10-108-260A-3561

Query Match 22.3%; Score 502; DB 15; Length 311;
Best Local Similarity 89.5%; Pred. No. 2.4e-31;
Matches 102; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 MKKMIRQFAIEYISKSGKTOENRNGSIGPSIVCKSIOMNQAEISLOEEOGGLDLTVNRM 60
DB 154 MKKMIRQFAIEYISKSGKTOENRNGSIGPSIVCKSIOMNQAEISLOEEOGGLDLTVNRM 213
QY 61 QEONTOQGGVLDLSTKTKTSIKSEESSICDPSSSENSVAGRLHNRNREDYVERSAE 114
DB 214 QEONTOQGGVLDLSTKTKTSIKSEESSICDPSSSENSVAGRLHNRNREDYVERSAE 267

RESULT 6

US-10-408-765A-1766
Sequence 1766, Application US/10408765A
Publication No. US20040101874A1
GENERAL INFORMATION:
APPLICANT: Ghosh, Soumitra S.
APPLICANT: Fahy, Eoin D.
APPLICANT: Zhang, Bing
APPLICANT: Gibson, Bradford W.
APPLICANT: Taylor, Steven W.
APPLICANT: Glenn, Gary M.
APPLICANT: Warnock, Dale E.
TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
FILE REFERENCE: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
FILE REFERENCE: 660088.465
CURRENT APPLICATION NUMBER: US/10/408,765A
CURRENT FILING DATE: 2003-04-04
NUMBER OF SEQ ID NOS: 3077
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1766
LENGTH: 572
TYPE: PRT
ORGANISM: Homo sapiens
US-10-408-765A-1766

Query Match 22.2%; Score 499.5; DB 16; Length 572;
Best Local Similarity 33.6%; Pred. No. 9.3e-31;
Matches 158; Conservative 66; Mismatches 143; Indels 103; Gaps 20;

QY 1 MKKMIRQFAIEYISKSGKTOE-----NRNGS-----IGPSIVCKSIOMNQAEISLOEE 48
DB 140 MQRMIOQFAAEYTSKNSSTQDPNSQSLPKASPVTTSTPAATTQNPVLSKLLMAD 199
QY 49 QSGPLDLTVNRQEQNTQGGVLDLSTKTKT-----SIKSEESSICDPSSSENSVAGRLHRN 104
DB 200 QDSPLDLTVRKQSEFSEQ-DGVLDLSTKSKPCAGSTLSHSPGCSSTQNGRPRPSQY 258
QY 105 REDYVERSAEFAADGLLSKALD-----IQSGALDINKAGILYGIQKTLILLLEALPAG 158
DB 259 RPD-----GLRSGDGVPPRLQDGTREGFGHSTSLKVPPLA-----RSLOISEELLRN 306
QY 159 K-----PASFPKNTKTRDFHDSYSYKDSKETCAVLQKVALWARAQAE-RTEKSKLN--- 206
DB 307 QLSTAASLGLSPGLQN-----HQGH-----LILSREASWAKPHVEFNLSRMKFRNG 352
QY 207 -LLETSEIKFPTASTYHLQTLQKMTQPKKNESLOYETSNPTVOLKIPOLRVSSVSKS 265
DB 353 ALSNISDLPLFAENSAPFQWALQ-----AKODGKKQVSHSSPVDLKIPOVRGMDLSWE 405
QY 266 QPDGSGLLDMYQVSKTSSVL-----EGSALQKLKNILPKQNKIEC--SGPVTHSSVDSYF 319
DB 406 SRTGD-----QYSYSLVMGSGQTESALSKLRLAILPKQSRKMLDADG-----DSWG 452
QY 320 LHGDLSPCLNSKNGTVDGTSNTDGLDRKDSQPKRKRGRYQYDHEIMEAIAVMMS 379
DB 453 SDAE-----QSTSGQPYTSDQEGD-----PGSKQPKRKRGRYQYNSILEAIVMS 502
QY 380 GKMSVSKAAGIYGVPHSTLEYKVKERSGTLTKPPKKLRL-----PDTGL 424
DB 503 GKMSVSKAAGIYGVPHSTLEYKVKERSGTLTKPPKKLRL-----PDTGL 424
RESULT 7
US-09-864-408A-2848

```

; Sequence 2848, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shinketsu, Richard A.
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Encod
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2848
; LENGTH: 104
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (24)..(24)
; OTHER INFORMATION: Wherein Xaa may be any naturally occurring amino acid
US-09-864-408A-2848

```

```

Query Match 13.1%; Score 294.5; DB 11; Length 104;
Best Local Similarity 74.4%; Pred. No. 1.3e-15;
Matches 58; Conservative 7; Mismatches 8; Indels 5; Gaps 1;

QY 352 SKQPKKRGYROYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLEYKVGSTLT 411
Db 7 SKQPKKRGYROYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLEYKVGSTLT 66

QY 412 PPKKKLRL-----PDGGL 424
Db 67 PPKKKLRLSEGPDSV 84

```

```

RESULT 8
US-10-016-768-2
; Sequence 2, Application US/10016768
; Publication No. US20020142443A1
; GENERAL INFORMATION:
; APPLICANT: Baehrecke, Eric H.
; TITLE OF INVENTION: GENES REGULATING PROGRAMMED CELL DEATH
; FILE REFERENCE: 4115-131
; CURRENT APPLICATION NUMBER: US/10/016,768
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(54)
; OTHER INFORMATION: X CAN BE ANY AMINO ACID
US-10-016-768-2

```

```

Query Match 12.1%; Score 273; DB 13; Length 53;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 353 KQPKKRGYROYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLEYKVK 405
Db 1 KQPKKRGYROYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLEYKVK 53

```

```

RESULT 9
US-10-016-768-3
; Sequence 3, Application US/10016768
; Publication No. US20020142443A1
; GENERAL INFORMATION:
; APPLICANT: Baehrecke, Eric H.

```

```

; TITLE OF INVENTION: GENES REGULATING PROGRAMMED CELL DEATH
; FILE REFERENCE: 4115-131
; CURRENT APPLICATION NUMBER: US/10/016,768
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 54
; TYPE: PRT
; ORGANISM: T. nigroviridis
US-10-016-768-3

```

```

Query Match 10.4%; Score 233.5; DB 13; Length 54;
Best Local Similarity 81.5%; Pred. No. 3.4e-11;
Matches 44; Conservative 7; Mismatches 2; Indels 1; Gaps 1;

QY 353 KQPKKRGYROYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLEYKVK 405
Db 1 KQPKKRGYROYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLEYKVK 54

```

```

RESULT 10
US-10-016-768-4
; Sequence 4, Application US/10016768
; Publication No. US20020142443A1
; GENERAL INFORMATION:
; APPLICANT: Baehrecke, Eric H.
; TITLE OF INVENTION: GENES REGULATING PROGRAMMED CELL DEATH
; FILE REFERENCE: 4115-131
; CURRENT APPLICATION NUMBER: US/10/016,768
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 53
; TYPE: PRT
; ORGANISM: M. musculus
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(54)
; OTHER INFORMATION: X can be any amino acid
US-10-016-768-4

```

```

Query Match 10.2%; Score 229; DB 13; Length 53;
Best Local Similarity 81.1%; Pred. No. 7.5e-11;
Matches 43; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 353 KQPKKRGYROYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLEYKVK 405
Db 1 KQPKKRGYROYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLEYKVK 53

```

```

RESULT 11
US-10-016-768-10
; Sequence 10, Application US/10016768
; Publication No. US20020142443A1
; GENERAL INFORMATION:
; APPLICANT: Baehrecke, Eric H.
; TITLE OF INVENTION: GENES REGULATING PROGRAMMED CELL DEATH
; FILE REFERENCE: 4115-131
; CURRENT APPLICATION NUMBER: US/10/016,768
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 1165
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-016-768-10

```

```

Query Match 8.9%; Score 200.5; DB 13; Length 1165;
Best Local Similarity 21.7%; Pred. No. 1.2e-06;
Matches 92; Conservative 66; Mismatches 146; Indels 119; Gaps 14;

```


RESULT 15
US-10-011-588-45
; Sequence 45, Application US/10011588
; Publication No. US20020168727A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Leonard
; TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
; TITLE OF INVENTION: NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
; TITLE OF INVENTION: RESEARCH AND CLINICAL THERAPY
; FILE REFERENCE: A34796 067252.0113
; CURRENT APPLICATION NUMBER: US/10/011,588
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: 09/910,186
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 09/611,419
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: 60/246,744
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: 60/311,966
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 848
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant protein encoded by SEQ ID NO:44
US-10-011-588-45

Query Match 5.9%; Score 132.5; DB 13; Length 848;
Best Local Similarity 20.4%; Pred. No. 0.18;
Matches 79; Conservative 64; Mismatches 129; Indels 115; Gaps 16;

Qy	25	GSIGPSIVCKSIQNAEQEQLDGLTNRMOEQNTQOGDGVLDLSTKTKSIKSE	84
Db	273	GGHDPSPVISTDMNINVKALQNFQD-----IANRLNVSSAOGSGI-DISLYKQIYKKNK	326
Qy	85	ESSICDPSSNSVAGRLHRNREDYVERSAEFADGLISKALKDIOGALDINKAGILYGP	144
Db	327	YDFVEDNGKYSV-----DKDKP-----DKLYKALMFGFTETNLAG-EYGI-	366
Qy	145	QKTLILLHL-EALP-----AGKPASPKNKTDFH-----DSYSYKDSKE	181
Db	367	-KTRYSYFSEYLPPIKTEKLLDNTIYTONEGFNIAKNLKTEFNGQNKAVNKEAYEISL	425
Qy	182	TCVLOKQVALWARAQARTKSKLNLLTSEIKFPPTASTYHLQTLTQKQWVTFQKRNESL	241
Db	426	EHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIAN-----KDSFSKDLAKAETI	477
Qy	242	QYETSNPTVQ-----LQIP-QLRVSSVSKSQP	267
Db	478	AYNTQNTNTIENFSIDQLILDNLSSGIDLFPNTEPFTNFDIDIDIEVIYKQSAKKKIFV	537
Qy	268	DGSGLLDVMYQVSKTSSVLEGSALQKLKNIPLKPKK-----IECSGPFVTHSSVDSY	318
Db	538	DGDSLFEYLHAQTFFSNI-ENQLTNSLNDALRNKKVYTFPTSTNLNVEKANTVVGAS-----	592
Qy	319	FLHGDLSPLCLNSKNGTVDG-TSENTE	344
Db	593	-----LFVNVKVGVIDDFTSESTQ	611

Search completed: February 24, 2005, 02:06:50
Job time : 132 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 24, 2005, 01:45:03 ; Search time 41 Seconds
(without alignments)
1037.264 Million cell updates/sec

Title: US-10-016-768A-8

Perfect score: 2250

Sequence: 1 MKKMTIRQFAIEYISKGKTQ.....GLYNWTDSTGCKNSKPKV 442

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79:*

1: Pirl:*

2: Pirl:*

3: Pirl:*

4: Pirl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	183	8.1	185	2 T24276	hypothetical prote
2	133.5	5.9	1109	2 A40801	phosphoprotein pho
3	131.5	5.8	997	2 T43523	cut17 protein - fi
4	131	5.8	1297	2 S39791	neurotoxin - clost
5	129.5	5.8	545	2 T19172	hypothetical prote
6	129.5	5.8	1560	2 T42727	proliferation pote
7	127	5.6	534	2 S61185	hypothetical prote
8	126.5	5.6	983	2 B49284	immediate-early pr
9	126.5	5.6	1078	2 T44232	hypothetical prote
10	126	5.6	433	2 T50395	actin-related prot
11	125.5	5.6	1819	2 A71928	cag island protein
12	125	5.6	948	2 T41496	conserved hypothet
13	124.5	5.5	2954	2 T14156	kinesin-related pr
14	124.5	5.5	3187	2 JC5837	364K Golgi complex
15	123.5	5.5	1827	2 T16270	hypothetical prote
16	123	5.5	1017	2 PC4035	cell-cycle-depende
17	123	5.5	1269	2 F84730	probable myosin he
18	122.5	5.4	958	2 T20621	hypothetical prote
19	122	5.4	911	2 S51441	hypothetical prote
20	122	5.4	1164	2 S46769	hypothetical prote
21	122	5.4	1435	1 BVBYL1	guanine nucleotide
22	121.5	5.4	1927	2 G64585	cag pathogenicity
23	121.5	5.4	2938	2 T30249	cell proliferation
24	121	5.4	833	2 T43446	hypothetical prote
25	120	5.3	1170	2 A56157	chromosome segrega
26	119	5.3	1148	2 A49651	replication factor
27	118.5	5.3	1313	2 P96673	hypothetical prote
28	118.5	5.3	2388	2 T14342	NSD1 protein - mou
29	118.5	5.3	5105	2 T32650	hypothetical prote

30	118	5.2	1147	2 JN0599	DNA-binding protei
31	118	5.2	1864	2 F86378	protein F21J9.12 {
32	117.5	5.2	2020	2 T21174	hypothetical prote
33	117	5.2	1146	2 A55532	myosin-heavy-chain
34	117	5.2	1802	2 S69703	HKR1 protein precu
35	117	5.2	1964	2 A59282	nonmuscle myosin I
36	116.5	5.2	1895	2 T19823	hypothetical prote
37	116.5	5.2	4350	2 T18440	hypothetical prote
38	116	5.2	696	2 T01951	hypothetical prote
39	116	5.2	1202	2 S55553	LAR-interacting pr
40	116	5.2	1272	2 C90593	hypothetical prote
41	115.5	5.1	1231	2 T18532	serine/threonine pr
42	115	5.1	927	2 T51536	hypothetical prote
43	115	5.1	1940	1 A24922	myosin heavy chain
44	114.5	5.1	651	2 A55100	SEC9 protein - yea
45	114.5	5.1	2845	2 I49505	adenomatous polyo

ALIGNMENTS

RESULT 1

T24276

hypothetical protein T01C1.3 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C;Accession: T24276

R;Lennard, N.

submitted to the EMBL Data Library, November 1995

A;Reference number: Z19868

A;Accession: T24276

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-185 <WIL>

A;Cross-references: UNIPROT:Q22051; EMBL:Z68010; PIDN:CAA92009.1; GSPDB:GN00028; CESP:T01

A;Experimental source: clone T01C1

C;Genetics:

A;Gene: CESP:T01C1.3

A;Map position: X

A;Introns: 25/3; 93/2; 131/3

C;Superfamily: Caenorhabditis elegans hypothetical protein T01C1.3

Query Match 8.1%; Score 183; DB 2; Length 185;

Best Local Similarity 32.4%; Pred. No. 0.00011;

Matches 47; Conservative 30; Mismatches 50; Indels 18; Gaps 5;

QY 263 SKSQPDGSLDVMYQVSKTSSVLEGSALQKL-KNILPKQNKIECSGPVTHSSVDVSYFLH 321

Db 9 TNSSEGTGETPMSD-KKSCSPLDPKWLESIWQNLFTQGNVV---PVDSANISNVDTTH 64

QY 322 GLDSPICLNSKNGTVDGTSNTDGLDRKDSKQPRKKRGYRQYDHEIMEEALAMVMSGK 381

Db 65 ---TPTPISEKQKHGNEE-----WKRSRPKRGYRKYDKNALDEAVRSVRGE 111

QY 382 MSVSKAQGIYGVPHSTLEYKVKERS 406

Db 112 MTVHRAGSFPGVPHSTLEYKVKERN 136

RESULT 2

A40801

phosphoprotein phosphatase (EC 3.1.3.16) 1 glycogen-binding regulatory chain - rabbit

N;Alternate names: protein phosphatase-1(G)

C;Species: Oryctolagus cuniculus (domestic rabbit)

C;Date: 10-Apr-1992 #sequence_revision 10-Apr-1992 #text_change 09-Jul-2004

C;Accession: A40801; S04004

R;Rang, P.M.; Bondor, J.A.; Swiderek, K.M.; DePaoli-Roach, A.A.

J. Biol. Chem. 266, 15782-15789, 1991

A;Title: Molecular cloning and expression of the regulatory (R-G1) subunit of the glycog

A;Reference number: A40801; MUID:91340717; PMID:1651919

A;Accession: A40801

A;Status: preliminary

A;Molecule type: mRNA

A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 932-997 <WED>
A;Cross-references: ENBL:AL121859; PIDN:CAB58376.1; GSPDB:GN00068; SPDB:SPCP31B10.10C
FEBS Lett. 248, 67-72, 1989
A;Title: Multisite phosphorylation of the glycogen-binding subunit of protein phosphatase
A;Reference number: S04004; MUID:89252053; PMID:2542090
A;Accession: S04004
A;Status: preliminary
A;Molecule type: protein
A;Residues: 33-68 <DEN>
C;Keywords: phosphoprotein; phosphoric monoester hydrolase

Query Match 5.9%; Score 133.5; DB 2; Length 1109;
Best Local Similarity 22.9%; Pred. No. 1.6; Indels 83; Gaps 17;
Matches 89; Conservative 53; Mismatches 163; Indels 83; Gaps 17;

QY 7 QFAIEYISKGTQENRNGSIGPVSICKSIQMNQAEISLQEBQGLDLTVNRMQEQNTQ 66
Db 208 EFCIRYETSVGTFSNNGT-NYTLVCQ-----KKEPEPPEPKPL-----BEAPSK 252

QY 67 QDGVLDLSTKTSIKSESSICDPSSNSVAGRLHNRNEDYVER---SAEFADGLLS-- 121
Db 253 QKKGCLKVKSSK-----ESS--ETSENNFEN--SKIADTVIPTIVCSHEEKEDLKSSY 303

QY 122 KALKDIQSGALDINKAGILYGIPOKTLHLHLALPAGK-----PASFKNKTDRDFHDSYS 175
Db 304 QNVKQVNTHEDEHNEKELELMINQRLIRTCASEYKGNLTSSDPNSIPNKPEELQKNQS 363

QY 176 YKDSKETCAVLQKVALWARAQARTKSKNLLLETSEIKFPTASTYHLQTLQKQVTPK 235
Db 364 H---SEACTDLSQRLSPGSAESSLKGFYHTE---KYSSGNESSHQ-----PSDMG 410

QY 236 EKNEISQYETSNPTVQLKIPOLRVSSVSKQPGSGLLDV-----MYQVSKTSSVLEGS 289
Db 411 EINPISLGGTSDGVSQVHLISSEKILDDNANPAHSGRGEISCSFPQQLKASNLKKYEGG 470

QY 290 ALQKLKNILPKQNKIECGSPVTHSSVDFLHGLDPLCLNSKNGVTVDGTSNTEDGLDR 349
Db 471 AENS-----EMKDCCLPRDVLKASDYF-----KKSTENRPSE--EDYGTS 510

QY 350 KDSKQPR-----KKGRYRYQYDHE 368
Db 511 KDNKEKRIQLDVDEKTSKNFRSIFYDQE 538

RESULT 3
T43523
cut17 protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 21-Jan-2000 #sequence revision 21-Jan-2000 #text_change 09-Jul-2004
C;Accession: T43523; T41649, T41700
R;Morishita, J.; Matsusaka, T.; Yanagida, M.
submitted to the EMBL Data Library, August 1999
A;Description: Fission yeast cut17 is required for chromosome segregation.
A;Reference number: Z22536
A;Accession: T43523
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-997 <MOR>
A;Cross-references: UNIPROT:O14064; EMBL:AB031034; PIDN:BAAB3415.1
R;Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, August 1998
A;Reference number: Z22007
A;Accession: T41649
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-997 <HAR>
A;Cross-references: EMBL:AL031323; PIDN:CAA20434.1; GSPDB:GN00068; SPDB:SPCC362.02C
A;Experimental source: strain 972h-; cosmid c962
A;Wedler, H.; Duesterhoef, A.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, October 1999
A;Reference number: Z22010
A;Accession: T41700

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 932-997 <WED>

A;Cross-references: ENBL:AL121859; PIDN:CAB58376.1; GSPDB:GN00068; SPDB:SPCP31B10.10C

C;Genetics:

A;Gene: cut17; SPCC962.02c; SPDB:SPCP31B10.10c

A;Map position: 3L

A;Introns: 43/3

Query Match 5.8%; Score 131.5; DB 2; Length 997;

Best Local Similarity 21.7%; Pred. No. 1.8; Indels 141; Gaps 24;

Matches 109; Conservative 77; Mismatches 175; Indels 141; Gaps 24;

QY 30 STVCSIQMNAENSLOEEQ-----EGPLDITVNRMQEQNTQQDGVLDLSTKTS 80

Db 408 SVVSKSEISSVSVSGKEQHTEKQVALETPEQQKVEKDEHLNLQG-SFIEESTKQ-P 465

QY 81 IKSESSICDPSSNSVAGRLHNRNEDYVERSAEFADGLLSK-----ALKD 126

Db 466 ISSKPESTSPDMTAAATGGRV-----SSSFRDKILQTNFSPRSTDSFSNISKRR 516

QY 127 IQSGALDIN-KAGILYGIPOKTLHLHLBALPAGKPAFPKN-----KTRDPHDSYSYKDSKE 181

Db 517 NSEERANDENETNLKIPKPKKQFQ-EVLQS-----KNILVSTSDSHEPVPKVTEDSQ 569

QY 182 TC AVLQKV-----ALWARAQARTKSKNLLLETSEIKFPTASTYHLQTLQK 229

Db 570 TAIHVSKPFEDLENKSESEQLLSSENDKPLDILPLILAIAK-----RKDN 618

QY 230 MYTQPEKNES-----LQYETSNPTVQLKIPOLRVSSV-----SKSQDPGSGLLDVMYQ- 278

Db 619 LVSGVLEKGGKSTSTKTFTDTSIVDF-TEKPKTEISEVLPEEKRAICADESQTVRVSDIR 677

QY 279 -VSKTSSVLEGSALQKLKNILPKQNKIECGSPVTHSSVDF-----HSSVD-----SY-- 318

Db 678 GVTKTRDVSSPVSDKSENV-----NHEANSIGHTVMVHSSLDPPQIVQPNLESGSYLK 733

QY 319 -----FLHGLDPLCLNSKNG-VTDGTSNTEDGLDRKDSKQPKKGRYR 363

Db 734 DLDPDRNVGSEKVTFTQEDDINSPLQSKNNQVAVNTETSDKLOEKA----- 782

QY 364 QYDHEI-----MEEAIAVMYSGKMSVSKAQGIYGVPHSTLEYKVKERSGTLKTPPKK-KL 417

Db 783 --NHELENIEKIEEKLTEV--DKVLSLDAFPQOEIKNSRTSVQNGTRSVSKNTPPEKTKV 838

QY 418 RLDPDTGLYNNMDSGTGCKNSS 439

Db 839 DKIDNVSKKDVETSPGSCETSS 860

RESULT 4

S39791

neurotoxin - Clostridium botulinum

C;Species: Clostridium botulinum

C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999

C;Accession: S39791

R;Campbell, K.; Collins, M.D.; East, A.K.

Biochim. Biophys. Acta 1216, 487-491, 1993

A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium a

A;Reference number: S39791; MUID:94032745; PMID:8268233

A;Accession: S39791

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1297 <CAM>

A;Cross-references: EMBL:X74162; NID:G441275; PIDN:CAAS2275.1; PID:G441276

C;Superfamily: tetanus toxin

C;Keywords: neurotoxin

Query Match 5.8%; Score 131; DB 2; Length 1297;

Best Local Similarity 20.2%; Pred. No. 2.7;

Matches 78; Conservative 64; Mismatches 131; Indels 114; Gaps 15;

QY 25 GSIGPSIVCKSIQMNQAEISLQEBQGLDLTVNRMQEQNTQQDGVLDLSTKTSIKSE 84

Db 273 GGHDPSPVSTDMYINAKLQNPQD-----IANRLNIVSSAOGSGI-DISLYKQYKXK 326
QY 85 ESSICDPSSNSVAGRLHRNRNEDVERSFAFADGLLSKALKIQSGALDINKAGILYIGIP 144
Db 327 YDFVPDNGKYSV-----DKDKF-----DKLYKALMFGFTETNLGAG-EVGI- 366
QY 145 OKTLLHLH-EALP-----AGKPASFNKTRDFH-----DSVSYKDSKE 191
Db 367 -KTRYSFSEYLLPIKTEKLLDNTIYQNEGFNIASKNLKTEFGNQKAVNKAYEISL 425
QY 182 TCALVOKVALWARAQARTSKNLNLETSEIKFPTASTYLHQLTQKMWTFQKKNESL 241
Db 426 EHLVIVRIACWKPWKYKNTGSKSQCIIVNNEDLFFIAN-----KDSFKDLAKAEYI 477
QY 242 QYETSNTPTQV-----LXIP-QLRVSVSSKQSP 267
Db 478 AYNTQNTIENFNSIDQLILDNDLSSGIDLPLNTEPFTNFDDIDIPVVIKQSAKAKIFV 537
QY 268 DSGGLDVMYQVSKTSVLSGSAQKLNILPKONK-----IECSGPVTHSSVDSY 318
Db 538 DGSLSFYLHAQTFPSNIENLQITNSLNDALRNNNKYTFPFTNLVEKANTVVGAS----- 593
QY 319 FLHGDLSPCLNSKNGTVDG-TSENTE 344
Db 594 -----LFVNVWKGVIDDFTSESTQ 612

RESULT 5
T19172
hypothetical protein F18C12.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T19172; T21088
R;Harris, B.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19083
A;Accession: T19172
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-545 <W12>
A;Cross-references: UNIPROT:O17584; EMBL:Z81466; PIDN:CA803870.1; GSPDB:GN00019; CESP:F18C12.3
A;Experimental source: clone C09H6
submitted to the EMBL Data Library, June 1996
A;Reference number: Z19371
A;Accession: T21088
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-545 <W12>
A;Cross-references: EMBL:Z75536; PIDN:CAA99833.1; GSPDB:GN00019; CESP:F18C12.3
A;Experimental source: clone F18C12
C;Genetics:
A;Gene: CESP:F18C12.3
A;Map position: 1
A;Introns: 171/3; 222/2; 316/3; 368/3; 409/1; 409/3; 493/1

Query Match 5.8%; Score 129.5; DB 2; Length 545;
Best Local Similarity 20.8%; Pred. No. 1.1;
Matches 93; Conservative 72; Mismatches 174; Indels 109; Gaps 20;

QY 2 KKMIRFA-----IEYISKGTQNRNGSIGPSIVCKSIQMNQASNSIQE---RQEGP 52
Db 175 KDSVAFKFAALNTKSLAELKINGTAEVAER--LMYKLVCSIIIVEKIQ-SIREKYTEMNKT 231
QY 53 LDLTVMRMOEQNTQQDGVLDLSTKTSI---KSESSICDPSSNSVAGRLHRNREDYV 109
Db 232 AAFQMRLGOQYAKPEKSKSSTKELNIENLDNNEDESEKEEIEENEEDYD 291
QY 110 ERSAEFADGLLSKALKXIQSGALDINKAGILYIGIPQKTLHLLEAL-----PAGKPSAF 163
Db 292 QSDIEMLDS-----DEBEAGEAAKNNRNNLLGLIGVQEDKNRPSLAPKR 337

QY 164 KXKTRDPHDSYKDSKETCAVLQKVALWARAQARTSKNLNLETSEIKFPTASTYLH 223
Db 338 KLVEEDEDVETVKNQKQKQIFKEP-----LKSKEVKNP----- 372
QY 224 QLTQKMWTFQKKNESLOVETSNPTVQLKIPOLRV--SSVSKSQPDGSGLLDVMYQVSK 281
Db 373 -----KSNKSTKKPKISAPIVKVKEKVEEENVSDDSDQKTLV-MKVDLSK 420
QY 282 TSVLGSAQKLNILPKONKIECSGPVTHSSVD-----SYPL--HGDLSPCLNSKNG 334
Db 421 GGIKAKA---QKFSSTTAPKSAKI--VAPVSEEDDDSSFFLPKSGVAVPRKIIPKK- 474
QY 335 TVDGTSENTEGLDKDSKQPKRGRYQYDHEIMEEALAMVMSCKMSVSKAQGIYGV 394
Db 475 ----PSENEK-VDRKFRQGTQKK-----SEAVVEKKKG-----SSKSAVSGEM 513
QY 395 H-STLEYKVKERSGTLKTPPKKLRPD 421
Db 514 HPSWIASQLKKELASAKPCGKKITFGD 541

RESULT 6
T42727
proliferation potential-related protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C;Accession: T42727
R;Witte, M.M.; Scott, R.E.
submitted to the EMBL Data Library, November 1998
A;Reference number: Z22246
A;Accession: T42727
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1560 <W12>
A;Cross-references: UNIPROT:P97868; EMBL:U83913; NID:g3858884; PID:g3858885; PIDN:AAC7243
A;Experimental source: strain Balb/C
C;Genetics:
A;Gene: P2P-R
C;Function:
A;Description: involved in hnRNP association and Rb1 binding
F;57-107/Domain: RING finger homology <RRN>

Query Match 5.8%; Score 129.5; DB 2; Length 1560;
Best Local Similarity 20.5%; Pred. No. 4.4;
Matches 103; Conservative 81; Mismatches 180; Indels 139; Gaps 23;

QY 15 KSGKTQENRNGSIGPSIVCKSIQMNQASNSIQEQQSGPLDLT-----VNRMOEQNTQQG 68
Db 786 KSDKTKRKSDG-----SATAKKONVLPKSGPQKVGDRKSPRPPPLKKAKEATK-- 839
QY 69 DGVLDLSTKTSIKSESSICDPSSNSVAGRLHRN---REDYVRSFAEFADGLLSKAL 124
Db 840 ----IDSVKPSSSSQKDEKVTGTPRAKHSKSAKEHQEAKPAKQKVKDC-----SKDI 899
QY 125 KIQSGALDINKAGILYIGIPQKTLHLLEALPAGKPSFNKTRDPHDSYKDSKETCA 184
Db 890 KSEKPSKD-EKA---KKPEKNKLIDSXGKKRKRKTEESKVDKDF-ESSSMKISKVEGT 943
QY 185 VLQKVALWARAQ-----ERT-EKSKNLNLETSEIKFPTASTYLHQLTQKMW-----VTQ 233
Db 944 EIVKPSPKKMGDVEKLEKTRPEKDKIASSTT-----PAKKTKLNRETGKKIGNAENAST 998
QY 234 FKKNESLOVETSNPTVQLKIPOLRVSSVSKSQPDGS-GLLDVMYQVSKTSVLSGSAQ 292
Db 999 TKEPSEKLESTSS-----KIKQEKVKAKRKVAGSESSSTLVDTYTSST-GGSPVR 1051
QY 293 KUK-----NILP-----KONKIECSGPVT 311
Db 1052 KSEKTDTRTVIKTMEYENNDNTAPAEVDIIMHVPSQKWDKDDPESEEDVKTTOPIQ 1111
QY 312 HSVDYSFLHGDLSPCLNSKNGTV--DGTSENTEGLDKDSKQPKRGRYQYDHEI 369
Db 1112 SVGKPSII-----KNVTKPSATAKYTE-----KESEQPEKQLQKPKASHL 1155

```

QY      370 MEEALAVMSGKMSVSKAOGIYGVPHSTLEVKYKERSGTLKTPKKKL----- 417
Db      1156 MQHEL-----RSKGSASSEKG-----RAKDRHSGSEKDNPDKKSKGAQPPKSTVD 1203

QY      418 RLPDTGLYNMTDSGTGSKCNSSK 440
Db      1204 RLSEQGHFKILSQSSKETRTSEK 1226

RESULT 7
S61185
N/Hypothetical protein YDR299w - yeast (Saccharomyces cerevisiae)
N/Alternate names: hypothetical protein D9740.7
C/Species: Saccharomyces cerevisiae
C/Date: 23-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 09-Jul-2004
C/Accession: S61185
R/Ding, H.
submitted to the EMBL Data Library, June 1995
A/Description: The sequence of S. cerevisiae cosmid 9740.
A/Reference number: S61160
A/Accession: S61185
A/Molecule type: DNA
A/Residues: 1-534 <DIN>
A/Cross-references: UNIPROT:Q06631; EMBL:U28374; NID:g849207; PID:g849214; GSPD:
C/Genetics:
A/Gene: SGD:BFR2; MIPS:YDR299w
A/Cross-references: SGD:S0002707
A/Map position: 4R

```

Query Match	5.6%	Score 127;	DB 2;	Length 534;	
Best Local Similarity	21.4%	Pred. No. 1.5;			
Matches	92;	Conservative 61;	Mismatches 157;	Indels 120;	Gaps 16
Qy	5	IRQFAIEYISKGTQENRNGSI-----GPSIVCKSIQMNQAEINSIQEQEGPLDITV	57		
Db	9	ISDIAIKPVNKFQDIEDENASLFOHNEKGES-----DLSDYGSNTEETKKAHYLEV	62		
Qy	58	NRMQEONTQQDGVLDLSTKK-TSIKSESSICDPSSSENSVAGRLHRRNPEDYVERSASF-	115		
Db	63	---EKSLRAEKGLGNLNDPKYTVGKSRQALYEEVSENEDEEEEEEKKEBDALSR	118		
Qy	116	-----ADG-----LILSKAL-----KDTQSGALDINKAGI	139		
Db	119	TDSEDEEVEIDEESDADGGETEEAAQKRHALSKLIQQETKQAINKLSQSVQDASKG-	176		
Qy	140	LYGIPOKTL-----LHLEALPAGKPASFKMTRDFHDSVSYKDSKETCAVLQKVALWA	193		
Db	177	-YSILQOTKLFNDNIIDLRIKQKAVIAANKUPLATTESWEEAKDDSEETKRLLK-----	229		
Qy	194	RAQARTEKSKNLNLETSEIKF-----PTASTYLHLTLQKMWTPQFKEKNESLQYETS	246		
Db	230	--ENEKLFNNLFRNLINFRIFKLPOLGDHITQNEEVAKHKLSSKLSKELYQETNSLDSELK	287		
Qy	247	N-PTVOLKIPQLRVSSVSKSQBDGSGL-----LDVMYQVKTSSVLEGSALQKLKN	296		
Db	288	EYRTAVLKNWSTKVSSASGNAALSNNKFKKAINLPADVQVENQLSDMSRLMGTGKTLNR-RN	346		
Qy	297	ILPKQNKTECS-----GPYTHSSVSQSYFLHGDLSPLCLNSKNGTVDGTSNTEDDGLD	348		
Db	347	ITPLYFQDCANGRLPELISFPVVKSVDD-----NENSDDGLD	384		
Qy	349	RKDSKQPRKK	358		
Db	385	IPKNYDPRRK	394		

RESULT 8
B49284
immediate-early protein RF3/RF4 - human herpesvirus 6 (strain Z29) (fragment)
C.Species: human herpesvirus 6
C.Date: 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 09-Jul-2004
C.Accession: B49284

R;Chou, S.; Marousek, G.I.
Virology 198, 370-376, 1994
A;Title: Analysis of interstrain variation in a putative immediate-early region of human
A;Reference number: A49284; MUID:94082474; PMID:8259673
A;Accession: B49284
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-983 <CHO>

Query Match	5.6%;	Score 126.5;	DB 2;	Length 983;
Best Local Similarity	18.7%;	Pred. No. 3.6;	Indels 81;	Gaps 13;
Matches	73;	Conservative	73;	Mismatches 164;
Qy	9	AIEYIKSGK-TQENRNGSIGPSIVCKSIQMNQAEKLSQEOEGPLDLTVNRMQEONTQO	67	
Db	642	AVSQCKSKKGTAKRKNVPKPS-KSKKIKLRLPET-----TNVIVISESEDEED	692	
Qy	68	GDGVLDLSTKTKTSIKRESSICDPSSSENSVAGLHRNREDYVERSAEFADGLLSKALKDI	127	
Db	693	GNNIIDKSMLEKTIKSE-----PNSESSSESDCTSEDNYLH-----LSDYDKVI	737	
Qy	128	QSGALDINKAGILYGPQKTLHLHLALPAGKPAFPKNKTRDPHDSYVSKDSKETCAVLQ	187	
Db	738	NNGHCQSK-----GFPSPVFTIPIKSMPC-----THDIRNKF-----VPK	772	
Qy	188	KVALWARAQARTE-----KSKLINLETSEIKFPTASTYVLHQLTLQKMWTPQKKNESL	241	
Db	773	KHVLWFMKTHKVDNCVITHSSAKMNVKNSDDVTEANHCFINHFVPKIKTDEEYKENVSY	832	
Qy	242	QY-----ETSNPVTOLKTPQLRVSSVSKSQPDGSGLLDMYQVSKTSVLEGSAL	291	
Db	833	TSYKIEDSKTDLEDITPTKLLITEMWENPMFDLTIKHGIACHQDLSKSYTVITHTAC	892	
Qy	292	QKLKNILPKNKIKTECSGPVTHSSVDSYFLHGLDSLPL-----CLNSKNGTVDGTSENT	343	
Db	893	EKNLVANSQNLVTAETQIIFDPQGT-----GNNSPILNIINDTTCQDENRCTEGTSNDN	947	
Qy	344	EDGLDRKDSKQPRKKGRYQY--DHEIMEE	372	
Db	948	EKCTIRSDCNSDKMEVEFKLDGYPSPDYDPFEE	978	

. RESULT 9
 T44232
 hypothetical protein U90 [imported] - human herpesvirus 6 (strain Z29)
 C:Species: human herpesvirus 6
 A:Variety: strain Z29
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
 C:Accession: T44232
 R:Dominguez, G.; Dambaugh, T.R.; Stamey, F.R.; Dewhurst, S.; Inoue, N.; Pallett, P.E.
 J. Virol. 73, 8040-8052, 1999
 A:Title: Human herpesvirus 6B genome sequence: coding content and comparison with human
 A:Reference number: Z22734; PMID:99412318; PMID:10482553
 A:Accession: T44232
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1078 <DOM>
 A:Cross-references: UNIPROT:Q9QJ15; EMBL:AF157706; PIDN:AAD49675.1
 A:Experimental source: strain Z29; variant B
 C:Genetics:
 A:Introns: 32/2; 103/1
 A:Note: U90

[illegible]

Db 788 GNIIIDKSMLEKTIKSE-----PNSSESSESDCTSEDNYLH-----LSDYDKVI 832
Qy 128 QSGALDINKAGILYGIPOKTLHLLEALPAGKPAFKNKTRDPHDSYVSKYDKSKETCAVLQ 187
Db 833 NNGHCQSK-----GFPSPVFTPIRSMGP-----THDIRNKF-----VPK 867
Qy 188 KVALWARAQAERTE-----KSKLNLETSEIKFPTASTYHLQTLQKMTQPKENESL 241
Db 868 KHLWLFMRKTHKVDNCVIHSSAKMKNKDSDVTEANHCFFINHEVPKTDDEEYERENVSY 927
Qy 242 QY-----ETSNPTVOLKIPQLRVSVSVSKSPQDGLDVMYQVSKTSSVLEGSAL 291
Db 928 TYSKIQDSKTDLEIPTTKKLIITEMWENEMDUTDIKIGIAKHQDLSKSYTVITHAC 987
Qy 292 QKLNILPKQNKIECSGPVTHSSVDSYFLHGLDLSPL-----CLNSKNGTVDGTSENT 343
Db 988 EKLANVANSQNLVTAETQIFDPQGT-----GNNSPILNIINDTTCQDNCRCTEGTSNDN 1042
Qy 344 EDGLDRKDSQPKRKRGRYQY--DHEIMEE 372
Db 1043 EKTIKSDCNDRKMEVFKLDGSDYDPPEE 1073

RESULT 10
T50395
actin-related protein [imported] - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jul-2004
R:Beck, A.; Borzym, K.; Reinhardt, R.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, January 1999
A:Reference number: Z25067
A:Accession: T50395
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-433 <BE>
A:Cross-references: UNIPROT:Q9P7X7; EMBL:AL136535; PIDN:CAR66436.1; GSPDB:GN00067; SPDB:
A:Experimental source: strain 972H(-); clone p1 p23A10
C:Genetics:
A:Gene: SPDB:SPBP23A10.08
A:Map position: 2
C:Superfamily: actin

Query Match 5.6%; Score 126; DB 2; Length 433;
Best Local Similarity 22.8%; Pred. No. 1.3;
Matches 95; Conservative 56; Mismatches 152; Indels 114; Gaps 19;
Qy 6 RQFAIEYISKGTQENRNGSIGPSIVCKSIQNNQAEV-----SLQEB-QSGPLD 54
Db 49 RLFGSEYIYKSNPGMEIKN-----AIRNGWVENMDVTVDLWRYGLEQQLKTNPLE 98
Qy 55 LTVNRMQEQNTQQGDVLDLSTKTSIK-----SSESSIC--DPSSENS 96
Db 99 HPILITEPDPNPNRVTLETWFESLRCPATYLAKQETCAAFASGKGTACLDVIGAES 158
Qy 97 VAGRLHNRREDYVERSAEAFADGLLSKALDIQSGAL-DINKAGILYGIPOKTLHLLEAL 155
Db 159 SVSAIY---DGFVLQGYQVQHSNAINDILAQTLDKN-----FEVMPKYLVSKNPV 210
Qy 156 PAKGPASFRNKRTRDFHDSYS-----YKDSKETCAVLQKVALWARAQAERTSKLNLL 208
Db 211 EIQGPANCLPRDITDYSVHQFQVRVYDEWKECALISDVPF-----SSETTI 259
Qy 209 ETSEIKFPTASTYL-----HOLTQKMTQFKEKNESLOYETSNPVTQ----- 251
Db 260 AESEFEFPDGRMWFAGRYQIPEHLFV---PGSDEEMNEEPSKPIEQETNENVSQDSS 316
Qy 252 -----LKIPQLRVSVSVSKSPD--GSGLLDVMYQVSKTSSVLSGSAQKLNILPK- 300
Db 317 VTNWTSRILGIPQLFQNCITSECVDVIRASLLNNVI--VCGGTSIMQGFSL-RLQNELSKL 373
Qy 301 ----QNKIECSGPVTHSSVDSYFLHGLDLSPLCLNSKNGTVD-----GTSENTEDGLDR 349

Db 374 YPGSRUKIHASGHVRSYASWLGSGILSSL-----GTFHQLWISRQRYEEHGSDR 424

RESULT 11
A71928
cag island protein - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Werberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric patho-
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: A71928
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1819 <ARN>
A:Cross-references: UNIPROT:Q9ZLV0; GB:AE001481; GB:AE001439; MID:g4155005; PIDN:AAD0604
A:Experimental source: strain J99
C:Genetics:
A:Gene: orf13/14

Query Match 5.6%; Score 125.5; DB 2; Length 1819;
Best Local Similarity 19.7%; Pred. No. 9.5;
Matches 102; Conservative 70; Mismatches 180; Indels 165; Gaps 18;
Qy 14 SKSGKTOENRNGSIGPSIVCKSIQNNQ-----AENSLOEQEGPLDITVNRMQEQNTQ 67
Db 72 SGNNETSESSNGSLADKLFKARKLVDDKRPPTQKSLDEETQ---KJNEEDQENNEHQ 128
Qy 68 GDGVLDLSTKTSIKSEESSICDPSSENSVAGRLHNRREDYVERSAEAFADGLLSKALKDI 127
Db 129 EETQTDLIDDETSEKTSQDQSPQLSNEATEA---NHFEDLLKESTESSDNHLN---- 180
Qy 128 QSGALDINKAGILYGIPOKTLHLLEALPAGKPAFKNK-----TRDFHDSYSYKD 178
Db 181 -----PTESSDNHLNDNPTETKTQETKTHFDEDKPEITDDSDQEIIG 224
Qy 179 SKE-----TCAVLQKVALWARA----- 195
Db 225 SKKKYIIGVAVLVILFIRSIFHYFVPLEDKSSRSKORNLVYNDEIQIRQEVNRL 284
Qy 196 QAERTKSKL-----NLLETSEI--KFPTASTY-----LHQL 225
Db 285 LKERNKGNMIDKNLFFNDPNRTLNYLNIAIEIDKNLRAFYECISNGGVEECKLI 344
Qy 226 TLQKMTQPKENESLOYETSNPVTQKIPQLRVSVSVSKSPQDGLDVMYQVSKTSSV 285
Db 345 KDKLQDQMKKTLEAYNDICKN-----AKTEBERIKCLDLIKDENLKKSL 389
Qy 286 LSGS-----ALQKLNILPKQNKIECSGPVTHSSVDSYF-----LHGLDLSPL--CLNS-- 331
Db 390 LNQKQVQVALDCLKNAKTDEERKECKLINDPEIRKFKLEQLQELQEKYKDCIKNAKT 449
Qy 332 ---XNGTVDGTSENTEDGLDR-----KDSQPKRKRGRYQYVDHEIMEEATA--MMSG 380
Db 450 EAEKNECKLGLSKEAIERLKQQAALDCLKNAKTDEERKECKLKNIPQDLQELLADMSVKAY 509
Qy 381 KMSVSKAQGIYGVPHSTLEYKVKERSGTLKTPPKKL 417
Db 510 KDCVSRAR-----NEKEQCEKELLTPEAKKL 536

RESULT 12
T41496
conserved hypothetical protein SPPC622.16c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
R:Seeger, K.; Harris, D.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.


```

Db      2378 TEALETIKESLSQKALDSFVKSMSLQDRDRIVSDYQLERHLSVILEKDEL--IQ 2435
Qy      296 NILPKQNKI--ECGSPVTHSSVDSYFLHGLDLSPLCLNSKNGTVDGT----SENTEDGLDR 349
Db      2436 DAAAEENKKEEIRG-----LRGHMDD--LNSNAKLDALIELQYRRDLNEVITI 2482
Qy      350 KDSKQPKKGRYQYDHEIMEBAIAMVMGSKMSVSKAQ 388
Db      2483 KDSQORQLLEAQLQQ-NKELNECVK--LEGRLKGSBAE 2518

RESULT 15
T16270
hypothetical protein F55D11.11 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T16270
R:Fulton, B.
submitted to the EMBL Data Library, June 1995
A:Description: The sequence of C. elegans cosmid F55D11.
A:Reference number: Z18487
A:Accession: T16270
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1827 <FUL>
A:Cross-references: UNIPROT:Q20042; EMBL:U29381; NID:g868214; PID:g868224; PIDN:AAA68757
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP.F55D11.11
A:Introns: 76/2; 131/3; 159/3; 185/3; 221/3; 253/3; 320/1; 369/3; 1133/3; 1205/2; 1250/1

Query Match      5.5%; Score 123.5; DB 2; Length 1827;
Best Local Similarity 19.6%; Pred. No. 13;
Matches 92; Conservative 71; Mismatches 206; Indels 101; Gaps 17;

Qy      18 KTOENRNGSIGPVSIVCKSIQMOAENSLQF--EQEGPLDLTVNM-----OEQNTQQG 68
Db      1028 QNSELKNKREG-----LSEKNNEERKKIQDLADQLREANKVHNMKMNLEEKNELD 1082
Qy      69 DGVLDLSTKTSIKSESSICDPSSNSVAGRLHRNREDYVERSABFADGLLSKALKDIQ 128
Db      1083 QNVTDLTNK--VRQLEIQMDKAANKNEVSGDLLRME-----HDAQSMLKQAQNE-Q 1131
Qy      129 SGALDINKAGILYGIPOKTLILLHLEALPAG---KPASFKNKTRDFHDSYSYKDSKETCAV 185
Db      1132 FRITDLEKVRKALQDENQRLVNDLATVKAFAEVKRETSKSAISDILDKYRSAEKANKGE 1191
Qy      186 LQKVAL-----WARAQARTF---KSKLNLETSEIKFPTASTYLHQL-----TLQKMV 231
Db      1192 LQNQLRSLDLATVTLKLEQELKAKDSNRLRDSOKRFEVQSKLANLQKSAVESLQNP 1251
Qy      232 TQPKENNESIQY-----ETSNPTVQLKIPQLRVSSVSXQPDGSGLLDVMYQV 279
Db      1252 SSNSRQNRSIYVDIPRAASSIGNENSDVPLRSSPSVRFADSSQNRQRAVDSMDVSSV 1311
Qy      280 SKTSSVLE-----GSALQKLKNIPLKQNKIECGPVTSHSSVDSYFLHGLDLS 326
Db      1312 GVTLRFLKERIEQLEADNADLSALEKAKDELQAKDELQAKDELQAKDELQAKDELQAKDEL 1366
Qy      327 LCINSKNGTVDG--TSN-----TEDGLDRKDSQPKRGRYQYDHEIMEBAIAMVMG 380
Db      1367 --ITEERNTIENRMTSQRYMLTNEESSRSREHRSRKARISTLEHLHREKESKLAHLR 1424
Qy      381 KMSVSKAQGIYGVPHSTLEYKVKERSGTLKTPPKKLLPLDPTGLYNMTDS 430
Db      1425 K-----EIEVLHGQHLDALESKEA-----TGLVGVDQS 1453

```

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 24, 2005, 01:33:28 ; Search time 171 Seconds
(without alignments)
1323.620 Million cell updates/sec

Title: US-10-016-768A-8

Perfect score: 2250

Sequence: 1 MKXMRQFAIEYISKSGTKQ.....GLYNMTDGTGCKNSKPKV 442

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2114.5	94.0	517	2 Q8CJG4	Q8CJG4 mus musculus
2	1090	48.4	213	2 Q96NKK1	Q96NKK1 homo sapien
3	1089	48.4	393	2 Q8C9J6	Q8C9J6 mus musculus
4	502	22.3	271	2 Q8N3X6	Q8N3X6 homo sapien
5	501.5	22.3	433	2 Q8BGT2	Q8BGT2 mus musculus
6	501.5	22.3	531	2 Q6ZPI3	Q6ZPI3 mus musculus
7	499.5	22.2	572	2 Q96JN0	Q96JN0 homo sapien
8	498	22.1	406	2 Q7Z723	Q7Z723 homo sapien
9	497.5	22.1	433	2 Q8N3L6	Q8N3L6 homo sapien
10	496.5	22.1	433	2 Q86T33	Q86T33 homo sapien
11	470	20.9	396	2 Q8C9Q0	Q8C9Q0 mus musculus
12	456.5	20.3	380	2 Q80VX8	Q80VX8 mus musculus
13	444	19.7	223	2 Q8C9B1	Q8C9B1 mus musculus
14	444	19.7	315	2 Q8BRN7	Q8BRN7 mus musculus
15	444	19.7	315	2 Q8BRT8	Q8BRT8 mus musculus
16	353.5	15.7	200	2 Q86T32	Q86T32 homo sapien
17	288	12.8	219	2 Q6NZG1	Q6NZG1 mus musculus
18	200.5	8.9	1165	2 Q9VDE0	Q9VDE0 drosophila
19	200.5	8.9	1188	2 Q7YU18	Q7YU18 drosophila
20	193.5	8.6	1598	2 Q95YM8	Q95YM8 apis mellif
21	183	8.1	185	2 Q22051	Q22051 caenorhabdi
22	141	6.3	689	2 Q9FNZ7	Q9FNZ7 oryza sativ
23	133.5	5.9	1109	1 P3JA_RABIT	Q00756 oryctolagus
24	133.5	5.9	2492	1 ATRX_PANTR	Q7YQM4 pan troglod
25	133.5	5.9	3616	2 Q9W6V0	Q9W6V0 gallus gall
26	131.5	5.8	997	1 BIR1_SCHPO	O14064 schizosacch
27	131.5	5.8	2492	1 ATRX_HUMAN	P46100 homo sapien
28	131	5.8	1296	1 BXG_CLOBO	O60393 clostridium
29	131	5.8	1489	2 Q63ZU6	Q63ZU6 xenopus lae
30	130.5	5.8	678	1 HOOK_DROVI	O61493 drosophila
31	130	5.8	1820	2 Q75XM5	Q75XM5 helicobacte

ALIGNMENTS

RESULT 1

ID	Q8CJG4	PRELIMINARY;	PRT;	517 AA.
AC	Q8CJG4;			
DT	01-MAR-2003 (Tremblrel. 23, Created)			
DT	01-MAR-2003 (Tremblrel. 23, Last sequence update)			
DT	01-MAR-2004 (Tremblrel. 26, Last annotation update)			
DE	Transcription factor MLR1.			
GN	Name=MLR1; Synonyms=mlr1;			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=22448616; PubMed=12560079; DOI=10.1016/S0014-5793(02)03858-9;			
RA	Kunieda T., Park J.M., Takeuchi H., Kubo T.;			
RT	"Identification and characterization of MLR1.2: two mouse homologues of Mblk-1, a transcription factor from the honeybee brain.";			
RL	FEBS Lett. 535:61-65(2003).			
DR	EMBL; AB076078; BAC20954.1; -			
DR	MED; MGI:2651932; Mlr1.			
DR	GO; GO:0005634; C:nucleus; IC.			
DR	GO; GO:0003702; P:RNA polymerase II transcription factor acti. . ; IDA.			
DR	GO; GO:006366; P:transcription from Pol II promoter; IDA.			
DR	InterPro; IPR007889; HTH_psq.			
DR	InterPro; IPR011526; HTH_psq_like.			
DR	Pfam; PF05225; HTH_psq; 1.			
DR	PROSITE; PS00960; HTH_PSQ; 1.			
SQ	SEQUENCE 517 AA; 57316 MW; C97403D3D296C52E CRC64;			
Query Match 94.0%; Score 2114.5; DB 2; Length 517;				
Best Local Similarity 94.1%; Pred. No. 9.1e-119;				
Matches 416; Conservative 12; Mismatches 13; Indels 1; Gaps 1;				
Qy	1	MKXMRQFAIEYISKSGTKQENRNGSIGPSIVCKSIQMNQANSLQEQSGPLDLTVNRM	60	
Db	77	MKXMRQFAIEYISKSGTKQENRNGSIGASLVCKSIQMNQADNCLQEQSGPLDLTVRT	136	
Qy	61	QEQVTOQGDVLDLSTKTKTSIKSESSICDPSSNSVAGRLHNRNEDYVERSAEFADGLL	120	
Db	137	QEQTAQGDVLDLSTKTKTSIKSESSISDPSSNSVAGRLHNRNEDYVERSAEFADGLL	196	
Qy	121	SKALKDIQSGALDINKAGILYGIPOKTLHLHALPAGKPAFPKTKTRDPHDSYSYKDSK	180	
Db	197	SKALKDIQSGALDINKAGILYGIPOKTLHLHALPAGKPAFPKTKTRDPHDSYSYKDSK	256	
Qy	181	ETCAVLQKVALWARAQARTKSKLNLETSEIKFPTASTYLHQLTLQKMWTFKKNES	240	
Db	257	ETCAVLQKVALWARAQARTKSKLNLETSEIKFPTASTYLHQLTLQKMWTFKKNES	316	
Qy	241	LQVETSNPTVLKIPOLRVSSVSKSPQDGSGLLDVMYQVSKTSVLEGSLQKLKILPK	300	
		: : : : : : : : : :		

017584 caenorhabdi
P97868 mus musculu
Q8mj06 papio hamad
Q86b12 dictyosteli
Q6cyg5 kluyveromyc
Q6vrs4 helicobacte
Q22126 caenorhabdi
Q06631 saccharomyc
Q66r08 saccharomyc
Q86kx8 dictyosteli
Q6ct81 kluyveromyc
Q69530 human herpe
Q6bhq0 debaryomyc
Q6fd39 acinetobact

32 129.5 5.8 545 2 017584
33 129.5 5.8 1591 2 P97868
34 129.5 5.8 2152 2 Q8MJ06
35 129.5 5.8 3099 2 Q86B12
36 129 5.7 1748 2 Q6CYG5
37 128 5.7 675 2 Q6VRS4
38 128 5.7 1256 2 Q22126
39 127 5.6 534 2 Q06631
40 127 5.6 534 2 Q66R08
41 127 5.6 1781 2 Q86KX8
42 126.5 5.6 951 2 Q6CT81
43 126.5 5.6 983 2 Q69530
44 126.5 5.6 1066 2 Q6BHGO
45 126.5 5.6 1073 2 Q6FD39

Matches 212; Conservative 1; Mismatches 0; Indels 0; Gaps 0

QY 230 MYTQPKKESLQVETSNPTVLQKIPQLRVSSVSKSQPDGSLDVMYQVSKTSSVLEGS 289
 Db 1 MYTQPKKESLQVETSNPTVLQKIPQLRVSSVSKSQPDGSLDVMYQVSKTSSVLEGS 60
 QY 290 ALQKLKNILPKQNKIECGPVTHSSVDSYFLHGLDLPCLNSKNGTVDGTSNTEDGLDR 349
 Db 61 ALQKLKNILPKQNKIECGPVTHSSVDSYFLHGLDLPCLNSKNGTVDGTSNTEDGLDR 120
 QY 350 KDSKQPRKRGYROYDHEIMEEALAMVSGKMSVSKAQGIYGVPHSTLEYKVKERSGTL 409
 Db 121 KDSKQPRKRGYROYDHEIMEEALAMVSGKMSVSKAQGIYGVPHSTLEYKVKERSGTL 180
 QY 410 KTPPKKKLRLEPDTGLYNNMTDGTGSCKNSSKPV 442
 Db 181 KTPPKKKLRLEPDTGLYNNMTDGTGSCKNSSKPV 213

RESULT 3
 Q8C9J6 PRELIMINARY; PRT; 393 AA.

AC Q8C9J6; 01-MAR-2003 (T-EMBLrel. 23, Created)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
 DE Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
 DE library, clone:A630051A18 product:hypothetical protein, full insert
 DE sequence. (Fragment).
 GN Name=Musli;
 OS Mus musculus (Mouse).
 OC Eukaryota; Euthera; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OK NCBI_TaxID=10090;
 RN [1]
 RP STRAIN=C57BL/6J; TISSUE=Thymus;
 RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RX Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA RIKEN FANTOM Consortium;
 RL "Functional annotation of a full-length mouse cDNA collection.";
 RN Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RX The FANTOM Consortium;
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RX MEDLINE=20530913; PubMed=1076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Konno H., Akizawa J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,

317 LQYETSNPPVQKIPQLRVSSVSKSQPDGSLDVMYQVSKTSSVLEGSALQKLKNILPK 376
 301 QNKIKCSGPVTHSSVDSYFLHGLDLPCLNSKNGTVDGTSNTEDGLDRKDSQPKKRG 360
 377 QNKLDGSGPVTHSSVDSYFLHGLDLPCLNSKNGTVDGTSNTEDGLDRKDNKQPKKRG 436
 361 RYROYDHEIMEEALAMVSGKMSVSKAQGIYGVPHSTLEYKVKERSGTLKTPPKKKLRLEP 420
 437 RYROYDHEIMEEALAMVSGKMSVSKAQGIYGVPHSTLEYKVKERSGTLKTPPKKKLRLEP 496
 421 DTGLYNNMTDGTGSCKNSSKPV 442
 497 DTGLY-MTDSGTGSCNNSKPV 517

RESULT 2
 Q96NKL PRELIMINARY; PRT; 213 AA.

AC Q96NKL; 01-DEC-2001 (T-EMBLrel. 19, Created)
 DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
 DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
 DE Hypothetical protein FLJ30696.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OK NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
 RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
 RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
 RA Sudo H., Hoshiro T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
 RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
 RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Chiba Y., Ishida S.,
 RA Tanai H., Kimata M., Watanabe M., Hirakawa S., Chiba Y., Ishida S.,
 RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
 RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
 RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
 RA Mueshino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
 RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
 RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
 RA Yamazaki M., Watanabe K., Kunigai A., Itakura S., Fukuzumi Y.,
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Omori Y., Okamoto S.,
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
 RA Tozashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
 RT "Complete sequencing and characterization of 21,243 full-length human
 RT cDNAs.";
 RL Nat. Genet. 36:40-45(2004).
 DR EMBL; AK055258; BAB70892.1; -;
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR InterPro; IPR007889; HTH_psq.
 DR Pfam; PF05225; HTH_psq; 1.
 DR PROSITE; PS00960; HTH_PSQ; 1.
 DR SEQUENCE 213 AA; 23477 MW; 4D7F6CABF95251B2 CRC64;

Query Match 48.4%; Score 1090; DB 2; Length 213;
 48.4%; Score 94; DB 2; Length 213;
 48.4%; Score 94; DB 2; Length 213;

RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Okamoto N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK041987; BAC31123.1; -.
DR MGD; MGI:2651932; Mlrl.
DR GO; GO:0005634; C:nucleus; IC.
DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. . .; IDA.
DR GO; GO:0006366; P:transcription from Pol II promoter; IDA.
KW Hypothetical protein.
FT NON TER 393
SQ SEQUENCE 393 AA; 43892 MW; 3742CF675978C6C3 CRC64;

Query Match 48.4%; Score 1089; DB 2; Length 393;
Best Local Similarity 92.3%; Pred. No. 2.1e-57;
Matches 216; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKKMIRQFAIEYISKSGTKQENRNGSIGPSIVCKSIQMNQAEISLOEEOGGLDLTVNRM 60
Db |||||
160 MKKMIRQFAIEYISKSGTKQENRNGSIGASVCKSIQMNQAEISLOEEOGGLDLTVNRM 219

QY 61 QEQTQQGGVLDLSTKTKTSIKSESSICDPSSSENSVAGRLHNRNREYVERSAEFADGLL 120
Db |||||
220 QEQTQQGGVLDLSTKTKTSIKSESSICDPSSSENSVAGRLHNRNREYVERSAEFADGLL 279

QY 121 SKALKDIQSGALDINKAGILYGIPTKLLHLLEALPAGPKASPKNKTRDFHDSYSYKDSK 180
Db |||||
280 SKALKDIQSGALDINKAGILYGIPTKLLHLLEALPAGPKASPKNKTRDFHDSYSYKDSK 339

QY 181 ETCVAVLQKVALWARAQAERTKSKLNLETSETKFTASTYHLHOLTQKMWTFQ 234
Db |||||
340 ETCVAVLQKVALWARAQAERTKSKLNLETSETKFTASTYHLHOLTQKMWTFQ 393

RESULT 4
Q8N3X6 PRELIMINARY; PRT; 271 AA.
AC Q8N3X6;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DE Hypothetical protein.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalek U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC037322; AAH37322.2; -.
KW Hypothetical protein.
SQ SEQUENCE 271 AA; 30284 MW; 0117DB9569556E032 CRC64;

Query Match 22.3%; Score 502; DB 2; Length 271;
Best Local Similarity 89.5%; Pred. No. 2.1e-22;
Matches 102; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 MKKMIRQFAIEYISKSGTKQENRNGSIGPSIVCKSIQMNQAEISLOEEOGGLDLTVNRM 60
Db |||||
114 MKKMIRQFAIEYISKSGTKQENRNGSIGPSIVCKSIQMNQAEISLOEEOGGLDLTVNRM 173

QY 61 QEQTQQGGVLDLSTKTKTSIKSESSICDPSSSENSVAGRLHNRNREYVERSAE 114
Db |||||
174 QEQTQQGGVLDLSTKTKTSIKSESSICDPSSSENSVAGRLHNRNREYVERSAE 227

RESULT 5
Q8BGT2 PRELIMINARY; PRT; 433 AA.
ID Q8BGT2;
AC Q8BGT2;
DT 01-WAR-2003 (TrEMBLrel. 23, Created)
DT 01-WAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Transcription factor MMR2 (Mus musculus adult male aorta and vein
cDNA, RIKEN full-length enriched library, clone:A530082C21
product:hypothetical protein, full insert sequence).
DE product:hypothetical protein, full insert sequence).
GN Name=Mmr2; Synonyms=mlr2;
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22448616; PubMed=12560079; DOI=10.1016/S0014-5793(02)03858-9;
RA Kunieda T., Park J.M., Takeuchi H., Kubo T.;
RT "Identification and characterization of Mmr1.2: two mouse homologues
of Mblk-1, a transcription factor from the honeybee brain.";
RL FEBS Lett. 535:61-65(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).

[4]
 RP STRAIN=C57BL/6J; TISSUE=Aorta and vein;
 RC THE FANTOM Consortium;
 RA THE RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs";
 RL Nature 420:563-573 (2002).
 [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes";
 RL Genome Res. 10:1617-1630 (2000).
 [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama S., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer";
 RL Genome Res. 10:1757-1771 (2000).
 [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
 RA Saito R., Saitoh H., Sakai K., Sakai K., Sakazume N., Sano H.,
 RA Saeki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB076079; BAC20955.1; -;
 DR EMBL; AK041090; BAC30816.1; -;
 DR MGD; MGI:2443930; Mlr2.
 DR GO; GO:0005634; C:nucleus; IC.
 DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. . . ; IDA.
 DR GO; GO:0006366; P:transcription from Pol II promoter; IDA.
 DR InterPro; IPR007889; HTH_psq.
 DR InterPro; IPR011526; HTH_psq_like.
 DR Pfam; PF05225; HTH_psq; 1.
 DR PROSITE; PS50960; HTH_PSQ; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 433 AA; 47124 MW; 736656D1F7E9A041 CRC64;
 Query Match 22.3%; Score 501.5; DB 2; Length 433;
 Best Local Similarity 34.5%; Pred. No. 4.3e-22;
 Matches 162; Conservative 62; Mismatches 143; Indels 103; Gaps 22;
 QY 1 MKKTIQPAIEYISKSGTQE-----NRNGS-----IGPSIVCKSIQMNQAEISLQEE 48
 DB 1 MORMIQOFAEYTSKTSSTQSPNKTQSLPKASPVTTSTAATQNPVLSKLLMAD 60
 QY 49 QEGPLDLTVNRMQEQNTQQGVDLSTKK-----TSIKSEESSICDPSSSENSVAGRLHN 104
 DB 61 QDSPLDLTVRKQSQSPSEQ-DGVLDLSTKKSPCASSTLSHSPGCSSTQGNRGRPSQY 119
 QY 105 REDIYVERSAEFADGLLSKALXD-----IQSGALDINKAGILYGIPOKTLHLLEALPAG 158
 DB 218 RPD-----GLRSGDGVPPRSLODGTREGFGHSTSLKVPLA-----RSLOISELLSRN 265

DB 120 RPD-----GLRSGDGVPPRSLODGTREGFGHSTSLKVPLA-----RSLOISELLSRN 167
 QY 159 K-----PASFKNKTDFHDSYSYKDSKTCVAVLQKVALWARAQAE-RTEKSKLN--- 206
 DB 168 QLSSTAASLSPGLQN-----HGQH-----LILSREASWAKPHYEFSLSRMKFRGNG 213
 QY 207 -LLBTSEIKFTASTYHLQLTLQKMTQFKKNESLQYETSNPTVOLKIPOLRVSSYSK 265
 DB 214 ALSNISDLPLAENS-----APFKVAHQTKQDGRK-DMSHSP-VDLKIPQVRGMOLSW 266
 QY 266 QPDGSLDLVMYQVSKTSSVL-----EGSALOKNLNLPKQNKIEC--SGPVTHSSVDSYF 319
 DB 267 SRTGD-----QVSYSVLVMGSGTESALSALLRAILPKQNKSMLDAGP-----DSWG 313
 QY 320 LHGDLSPCLNSKNGTVDGTSNTEDGLDRKDSQPKRKRGRYQYDHHEIMEEIAIAMS 379
 DB 314 SDAE-----QSTGQPYPTSDQEGD-----PGSKPRKRGRYQYNSIIEEIAISV 363
 QY 380 GRMSVSKAQGIYGVPHSTLEYKVKERSGTILTPPKKLRL-----PDTGL 424
 DB 364 GRMSVSKAQSIYGVPHSTLEYKVKERLGTLPKPPKQKMLMRSEGFVSV 413
 RESULT 6
 Q6ZPI3 PRELIMINARY; PRT; 531 AA.
 AC Q6ZPI3;
 DT 05-JUL-2004 (Tremblrel. 27, Created)
 DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
 DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
 DE MKIAA1795 protein (Fragment).
 GN Name=mkIAA1795;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC TISSUE=Embryonic tail;
 RX PubMed=14621295;
 RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S.,
 RA Saga Y., Nagase T., Ohara O., Koga H.;
 RT Prediction of the coding sequences of mouse homologues of KIAA gene:
 RT III. the complete nucleotide sequences of 500 mouse KIAA-homologous
 RT cDNAs identified by screening of terminal sequences of cDNA clones
 RT randomly sampled from size-fractionated libraries.";
 RL DNA Res. 10:167-180 (2003).
 DR EMBL; AK129442; BAC98252.1; -;
 DR GO; GO:0005634; C:nucleus; IC.
 DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. . . ; IDA.
 DR GO; GO:0006366; P:transcription from Pol II promoter; IDA.
 DR InterPro; IPR007889; HTH_psq.
 DR InterPro; IPR011526; HTH_psq_like.
 DR Pfam; PF05225; HTH_psq; 1.
 DR PROSITE; PS50960; HTH_PSQ; 1.
 FT NON_TER 1
 SQ SEQUENCE 531 AA; 58235 MW; EA97247C413A403B CRC64;
 Query Match 22.3%; Score 501.5; DB 2; Length 531;
 Best Local Similarity 34.5%; Pred. No. 5.7e-22;
 Matches 162; Conservative 62; Mismatches 143; Indels 103; Gaps 22;
 QY 1 MKKTIQPAIEYISKSGTQE-----NRNGS-----IGPSIVCKSIQMNQAEISLQEE 48
 DB 99 MORMIQOFAEYTSKTSSTQSPNKTQSLPKASPVTTSTAATQNPVLSKLLMAD 158
 QY 49 QEGPLDLTVNRMQEQNTQQGVDLSTKK-----TSIKSEESSICDPSSSENSVAGRLHN 104
 DB 159 QDSPLDLTVRKQSQSPSEQ-DGVLDLSTKKSPCASSTLSHSPGCSSTQGNRGRPSQY 217
 QY 105 REDIYVERSAEFADGLLSKALXD-----IQSGALDINKAGILYGIPOKTLHLLEALPAG 158
 DB 218 RPD-----GLRSGDGVPPRSLODGTREGFGHSTSLKVPLA-----RSLOISELLSRN 265

Qy	207	-LLETSEIKFPTASTYIHLQTLQKMTQFKENKESLQVETSNPTVOLKIPLQRLSVSSVSKS	266
Db	353	ALSNISDLPLFLAENSAPFKMALQ-----AKODGKKDVSHSSPYDLKXIPQVRGMDLSWE	405
Qy	266	QPGSGLLDVMYQVSKTSSVL-----EGSALOKLKNILPKONKIEC--SGPVTHSSVDVSF	319
Db	406	SRTGD-----QVSYSLSVWGSQTESALSKKLRAILPRQKSKMLDAGP-----DSWG	452
Qy	320	LHGDSLPLCLNSKNGTVDTGSENTEDGLDRKDSQPRKKRGYRYDHEIMEEAIAMVMS	379
Db	453	SDAE-----QSTSGQPYPTSDQBG---PGSQPRKKRGYRYNSIIEEALISVMS	502
Qy	380	GKMSVSKAQSIYGVPHSTLEYKVKVERSGTLTKTPKKLURL-----PDTGL	424
Db	503	GKMSVSKAQSIYGIPIHSTLEYKVKERLGTGLKNPPKKKKMLMRSEGPDSV	552

RESULT 8	
Q7Z723	PRELIMINARY; PRT; 406 AA.
ID	Q7Z723
AC	Q7Z723;
DT	01-OCT-2003 (TrEMBLrel. 25, Created)
DT	01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE	MUR2 protein.
OS	Homo sapiens (Human)
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC	NCBI_TaxID=9606;
RP	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Lung;
RX	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA	Datchenko L., Marusina K., Farmer A.A., Rubin G.W., Hong L.,
RA	Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA	Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA	Richards S., Worley K.C., Haile S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA	Faney J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA	Blakeley R.W., Touchman J.W., Green E.D., Dickens M.C.,
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA	Krzywinski M.I., Skaleka U., Smalusz D.E., Schnerch A., Schein J.E.,
RA	Jones S.J., Marra M.A.;
RT	"Generation and initial analysis of more than 15,000 full-length human
RT	and mouse cDNA sequences."
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RP	[2]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Lung;
RL	Strausberg R.;
DR	Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR	EMBL; BC053359; AAH53359.1; -
DR	GO; GO:0005634; C:nucleus; IEA.
DR	GO; GO:0003677; F:DNA binding; IEA.
DR	InterPro; IPR007889; HTH_psq
DR	InterPro; IPR011526; HTH_psq_like.
DR	Pfam; PF05225; HTH_psq; 1.
DR	PROSITE; PS00960; HTH_PSQ; 1.
SQ	SEQUENCE 406 AA; 44083 MW; ACE760CF3CB1F2EF CRC64;

Query Match	22.1%;	Score 498;	DB 2;	Length 406;
Best Local Similarity	33.98;	Pred. No. 6.4e-22;		
Matches 156;	Conservative 65;	Mismatches 141;	Indels 98;	Gaps 19

Qy	1	MKKWIROFALEYTSKSGKTQE-----NRNGS-----IGPSIVCKSIQMNAENSLQGE	48
----	---	---	----

```
Db 1 MORMIQFAAEYTSKNSSTQDPSPNSTKNQSLPKASPVTTSTPTAATQNPVLSKLLMAD 60
QY 49 QEGPLDLTVNRMOEQNTQCGDGLDLSTKKT-----SIKSEESSICDPSSNSVAGRLHRN 104
Db 61 QDSPLDLTVRKSSQSEPSQ-DGVLDLSTKSPCAGSTSLSHSPGCCSTQGNRPGPSQY 119
QY 105 REDYVERSAEFAADGLLSKALKD-----IQSGALDINKAGILYIGIPQKTLILLHLEALPAG 158
Db 120 RPD-----GLRSGDGVPPRSIQDGTREGFGHSTSLKVPLA-----RSIQISELLSRN 167
QY 159 K-----PASFKNKTRDFHDSYSYKDSKETCAVLQKVALWARAQAE-RTEKSKLN--- 206
Db 168 QLSTAASLSPGSLQN-----HGQH-----LILSREASWAKPHYFNLSRMKFRNG 213
QY 207 -LLETSEIKFPTASTYHLQTLQKMTQPKKNESQYETSNTVOLKIPOLRVSSVSKS 265
Db 214 ALSNISDLPLAENSAPFKWALQ-----AKQDGKDVSHSPVDLKIPIQVRGMDLSWE 266
QY 159 K-----PASFKNKTRDFHDSYSYKDSKETCAVLQKVALWARAQAE-RTEKSKLN--- 206
Db 168 QLSTAASLSPGSLQN-----HGQH-----LILSREASWAKPHYFNLSRMKFRNG 213
QY 207 -LLETSEIKFPTASTYHLQTLQKMTQPKKNESQYETSNTVOLKIPOLRVSSVSKS 265
Db 214 ALSNISDLPLAENSAPFKWALQ-----AKQDGKDVSHSPVDLKIPIQVRGMDLSWE 266
QY 266 QPDGSGLLDWMYQVSKTSSVL-----EGSALQKLKNILPKONKIEC--SGPVTHSSVDSYF 319
Db 267 SRTGD-----QYSYSSLVMGQTESALSCLKRALILPKSRKSMLDAGP-----DSWG 313
QY 320 LHGDLSPLCLNSKNGTVDGTSNTEDGLDRKQKPRKGRYQYDHEIMEEAIAMWS 379
Db 314 SDAEQS-----QSTSGOPYPTSDQEGD-----PGSKQPKRKGRYQYNSILEEAISSVMS 363
QY 380 GKMSVSKAQGIYGVPHSTLEYKVKERSGTLKTPPKKKLRL 419
Db 364 GKMSVSKAQSIYGIPIHSTLEYKVKERLGTILKNPPKKMKML 403

RESULT 9
Q8N3L6 PRELIMINARY; PRT; 433 AA.
ID Q8N3L6 Q8N3L6
AC Q8N3L6 Q8N3L6
DT 01-OCT-2002 (T-REMBLrel. 22, Created)
DT 01-JUN-2003 (T-REMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (T-REMBLrel. 26, Last annotation update)
DE Hypothetical protein DKFZp451A142.
GN Name=DKFZp451A142;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Human skeletal muscle;
RA Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wiemann S.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL834245; CAD38921.2; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR007889; HTH_psq.
DR InterPro; IPR011526; HTH_psq_like.
DR Pfam; PF05225; HTH_psq; 1.
DR PROSITE; PS50960; HTH_PSQ; 1.
KW Hypothetical protein.
SQ SEQUENCE 433 AA; 47016 MW; 9EDA693DEAF7AB64 CRC64;

Query Match 22.1%; Score 497.5; DB 2; Length 433;
Best Local Similarity 33.4%; Pred. No. 7.4e-22;
Matches 157; Conservative 64; Mismatches 146; Indels 103; Gaps 19;

QY 1 MORMIQFAAEYTSKNSSTQDPSPNSTKNQSLPKASPVTTSTPTAATQNPVLSKLLMAD 60
Db 1 MORMIQFAAEYTSKNSSTQDPSPNSTKNQSLPKASPVTTSTPTAATQNPVLSKLLMAD 60
QY 49 QEGPLDLTVNRMOEQNTQCGDGLDLSTKKT-----SIKSEESSICDPSSNSVAGRLHRN 104
Db 1 MORMIQFAAEYTSKNSSTQDPSPNSTKNQSLPKASPVTTSTPTAATQNPVLSKLLMAD 60
QY 49 QEGPLDLTVNRMOEQNTQCGDGLDLSTKKT-----SIKSEESSICDPSSNSVAGRLHRN 104
Db 61 QDSPLDLTVRKSSQSEPSQ-DGVLDLSTKSPCAGSTSLSHSPGCCSTQGNRPGPSQY 119
QY 105 REDYVERSAEFAADGLLSKALKD-----IQSGALDINKAGILYIGIPQKTLILLHLEALPAG 158
Db 120 RPD-----GLRSGDGVPPRSIQDGTREGFGHSTSLKVPLA-----RSIQISELLSRN 167
QY 159 K-----PASFKNKTRDFHDSYSYKDSKETCAVLQKVALWARAQAE-RTEKSKLN--- 206
Db 168 QLSTAASLSPGSLQN-----HGQH-----LILSREASWAKPHYFNLSRMKFRNG 213
QY 207 -LLETSEIKFPTASTYHLQTLQKMTQPKKNESQYETSNTVOLKIPOLRVSSVSKS 265
Db 214 ALSNISDLPLAENSAPFKWALQ-----AKQDGKDVSHSPVDLKIPIQVRGMDLSWE 266
QY 266 QPDGSGLLDWMYQVSKTSSVL-----EGSALQKLKNILPKONKIEC--SGPVTHSSVDSYF 319
Db 267 SRTGD-----QYSYSSLVMGQTESALSCLKRALILPKSRKSMLDAGP-----DSWG 313
QY 320 LHGDLSPLCLNSKNGTVDGTSNTEDGLDRKQKPRKGRYQYDHEIMEEAIAMWS 379
Db 314 SDAEQS-----QSTSGOPYPTSDQEGD-----PGSKQPKRKGRYQYNSILEEAISSVMS 363
QY 380 GKMSVSKAQGIYGVPHSTLEYKVKERSGTLKTPPKKKLRL 419
Db 364 GKMSVSKAQSIYGIPIHSTLEYKVKERLGTILKNPPKKMKML 403
```

```
QY 105 REDYVERSAEFAADGLLSKALKD-----IQSGALDINKAGILYIGIPQKTLILLHLEALPAG 158
Db 120 RPD-----GLRSGDGVPPRSIQDGTREGFGHSTSLKVPLA-----RSIQISELLSRN 167
QY 159 K-----PASFKNKTRDFHDSYSYKDSKETCAVLQKVALWARAQAE-RTEKSKLN--- 206
Db 168 QLSTAASLSPGSLQN-----HGQH-----LILSREASWAKPHYFNLSRMKFRNG 213
QY 207 -LLETSEIKFPTASTYHLQTLQKMTQPKKNESQYETSNTVOLKIPOLRVSSVSKS 265
Db 214 ALSNISDLPLAENSAPFKWALQ-----AKQDGKDVSHSPVDLKIPIQVRGMDLSWE 266
QY 266 QPDGSGLLDWMYQVSKTSSVL-----EGSALQKLKNILPKONKIEC--SGPVTHSSVDSYF 319
Db 267 SRTGD-----QYSYSSLVMGQTESALSCLKRALILPKSRKSMLDAGP-----DSWG 313
QY 320 LHGDLSPLCLNSKNGTVDGTSNTEDGLDRKQKPRKGRYQYDHEIMEEAIAMWS 379
Db 314 SDAEQS-----TPGOPYPTSDQEGDPSKQPKRKGRYQYNSILEEAISSVMS 363
QY 380 GKMSVSKAQGIYGVPHSTLEYKVKERSGTLKTPPKKKLRL-----PDTGL 424
Db 364 GKMSVSKAQSIYGIPIHSTLEYKVKERLGTILKNPPKKMKMLRSEGEDVSV 413

RESULT 10
Q86T33 PRELIMINARY; PRT; 433 AA.
ID Q86T33 Q86T33
AC Q86T33 Q86T33
DT 01-JUN-2003 (T-REMBLrel. 24, Created)
DT 01-JUN-2003 (T-REMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (T-REMBLrel. 26, Last annotation update)
DE Hypothetical protein DKFZp313C088.
GN Name=DKFZp313C088;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RG The German cDNA Consortium;
RA Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wiemann S.;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL832106; CAD91159.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR007889; HTH_psq.
DR InterPro; IPR011526; HTH_psq_like.
DR Pfam; PF05225; HTH_psq; 1.
DR PROSITE; PS50960; HTH_PSQ; 1.
KW Hypothetical protein.
SQ SEQUENCE 433 AA; 46975 MW; 9089929710DF291E CRC64;

Query Match 22.1%; Score 496.5; DB 2; Length 433;
Best Local Similarity 33.6%; Pred. No. 8.5e-22;
Matches 158; Conservative 65; Mismatches 144; Indels 103; Gaps 20;

QY 1 MORMIQFAAEYTSKNSSTQDPSPNSTKNQSLPKASPVTTSTPTAATQNPVLSKLLMAD 60
Db 1 MORMIQFAAEYTSKNSSTQDPSPNSTKNQSLPKASPVTTSTPTAATQNPVLSKLLMAD 60
QY 49 QEGPLDLTVNRMOEQNTQCGDGLDLSTKKT-----SIKSEESSICDPSSNSVAGRLHRN 104
Db 61 QDSPLDLTVRKSSQSEPSQ-DGVLDLSTKSPCAGSTSLSHSPGCCSTQGNRPGPSQY 119
QY 105 REDYVERSAEFAADGLLSKALKD-----IQSGALDINKAGILYIGIPQKTLILLHLEALPAG 158
Db 120 RPD-----GLRSGDGVPPRSIQDGTREGFGHSTSLKVPLA-----RSIQISELLSRN 167
QY 159 K-----PASFKNKTRDFHDSYSYKDSKETCAVLQKVALWARAQAE-RTEKSKLN--- 206
```

```

Db 168 QLSTAAALGSPGLQ-----HGQH-----LILSRASWAKPHYFNLNRKFRGNG 213
Qy 207 -LLETSEIKPPTASTYLHQLTQMWVTFQKEKNESLQYETSNPTVLKIPQLRVSYSVKS 265
Db 214 ALSNISDLPLAENSAPPKMALQ-----AKDGKDVSHSHSPVDLKIPQVRGMDLSWE 266
Qy 266 QPDGSGLLDMVQVSKTSSVL-----EGSALQKLKNILPKQNKIEC--SGPVTHSSVDSYF 319
Db 267 SRTGD-----QYSYSSLVMGQTESALSKKLRAILPKQSRKSMLDAGP-----DSWG 313
Qy 320 LHGDLSPLCLNSKNGTVDTGTSNTDGLDRKDSQPKRKRGRVQYDHEIMEBAIAVMMS 379
Db 314 SDAE-----QSTSGQPYPTSDQEGD-----PGSKQPKRKRGRVQYNSSEILEBAISVMS 363
Qy 380 GRMSVSKAQGIYGVPHSTLTYKVKERSGTLKTPPKKRL-----PDTGL 424
Db 364 GRMSVSKAQSIYGIHSTLTYKVKERLGLTKNPKPKQKMLRSEGGPDVSV 413

RESULT 11
Q8C9Q0 PRELIMINARY; PRT; 396 AA.
AC Q8C9Q0;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DE Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
DE library, clone:A630025C20 product:hypothetical protein, full insert
DE sequence. (Fragment).
GN Name=Mlr2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA The FANTOM Consortium.
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,

```

```

RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tomaru A., Toyota T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK041621; BAC31007.1; --
DR MGD; MGI:2443930; Mlr2.
DR GO; GO:0005634; C:nucleus; IC.
DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. .; IDA.
DR GO; GO:0006366; P:transcription from Pol II promoter; IDA.
DR InterPro; IPR007889; HTH_psq.
DR InterPro; IPR011526; HTH_psq_like.
DR Pfam; PF05225; HTH_psq; 1.
DR PROSITE; PS0960; HTH_PSQ; 1.
KW Hypothetical protein.
FT NON_TER 396
FT SEQUENCE 396 AA; 43085 MW; EE4A585F62336B35 CRC64;
Qy 1 MKKMIQFAIEYISKSGKQEQ-----NRNGS-----IGPSIVCKSIQNNQAEISQEE 48
Db 1 MQEMIQFAAEYTSKTSSTQDPQNPSTKNQSLPKASPVTTSPTAATQNPVLSKLLMAD 60
Qy 49 QEGFLDTVNRMQEONTQGDGVLDLSTKK-----TSIKSEESICDPSSNSVAGRLHRN 104
Db 61 QDSPLDLTVRKSSQSEPEQ-DGVLDLSTKSPCASSTLSHSPGCSSTQNGRPGRPQSY 119
Qy 105 REDYVERSAEFADGLLSKALD-----IQSGALDINKAGILYIPQKTLHLLEALPAG 158
Db 120 RPD-----GLRSGDGVPPRLSDQGTREGFGHSTLSKVPLA-----RSIQSEELSRN 167
Qy 159 K-----PAGFKNKTTRDPHDSYSYKDSKETCAVLQKQVALWARAQAE-RTEKSKLN--- 206
Db 168 QLSTAAALGSPGLQ-----HGQH-----LILSRASWAKPHYFNLNRKFRGNG 213
Qy 207 -LLETSEIKPPTASTYLHQLTQMWVTFQKEKNESLQYETSNPTVLKIPQLRVSYSVKS 265
Db 214 ALSNISDLPLAENS-----APPKMAHQTKDQGR-DMSHSSP-VDLKIPQVRGMDLSWE 266
Qy 266 QPDGSGLLDMVQVSKTSSVL-----EGSALQKLKNILPKQNKIEC--SGPVTHSSVDSYF 319
Db 267 SRTGD-----QYSYSSLVMGQTESALSKKLRAILPKQSRKSMLDAGP-----DSWG 313
Qy 320 LHGDLSPLCLNSKNGTVDTGTSNTDGLDRKDSQPKRKRGRVQYDHEIMEBAIAVMMS 379
Db 314 SDAE-----QSTSGQPYPTSDQEGD-----PGSKQPKRKRGRVQYNSSEILEBAISVMS 363
Qy 380 GRMSVSKAQGIYGVPHSTLTYKVKERSGTLKTP 412
Db 364 GRMSVSKAQSIYGIHSTLTYKVKERLGLTKNP 396

RESULT 12
Q8OVA8

```

Q80VAB	PRELIMINARY;	PRT;	380 AA.
Q80VAB;			
DT 01-JUN-2003	(T-EMBLrel. 24, Created)		
DT 01-JUN-2003	(T-EMBLrel. 24, Last sequence update)		
DT 01-MAR-2004	(T-EMBLrel. 26, Last annotation update)		
DE	M12 protein (Fragment).		
GN	Name=M12;		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=FVB/N; TISSUE=Mammary tumor;		
RX	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;		
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,		
RA	Klauser R.D., Collins F.S., Wagner K.H., Shenmen C.M., Schuler G.D.,		
RA	Altschul S.F., Zebberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,		
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,		
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,		
RA	Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,		
RA	Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,		
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,		
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,		
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,		
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,		
RA	Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,		
RA	Whiting M., Madan A., Young A.C., Shcherchenko Y., Bouffard G.G.,		
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,		
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,		
RA	Krzywicki M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,		
RA	Jones S.J., Marra M.A.		
RT	"Generation and initial analysis of more than 15,000 full-length human		
RT	and mouse cDNA sequences."		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=FVB/N; TISSUE=Mammary tumor;		
RA	Strausberg R.;		
RL	Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BC050068; AAH50068.1; -		
DR	MGD; MGI:2443930; M12.		
DR	GO; GO:0005634; C:nucleus; IC.		
DR	GO; GO:0003702; F:RNA polymerase II transcription factor acti. . ; IDA.		
DR	GO; GO:0006366; P:transcription from Pol II promoter; IDA.		
DR	InterPro; IPR007889; HTH_psq.		
DR	InterPro; IPR011526; HTH_psq_like.		
DR	Pfam; PF05225; HTH_psq; 1.		
DR	PROSITE; PS50960; HTH_PSQ; 1.		
FT	NON TER		
SQ	SEQUENCE 380 AA; 41445 MW; 4D3B6820C3B15EEE CRC64;		
Query Match	20.3%; Score 456.5; DB 2; Length 380;		
Best Local Similarity	35.3%; Pred. No. 1.8e-19;		
Matches 146;	Conservative 54; Mismatches 123; Indels 91; Gaps 20;		
QY	45 LQEEGRLDITVNRMQEQTQGGDGLDITSTKK----TSIKSEESICDPSSENSVAGR 100		
DB	4 LMDQSDPLDITVRKSKQSEFSEQ-DGVLDTLTKSPCASTSLSHSPGCSSTGNGRPGR 62		
QY	101 LHRNREDYVERSAEFADGLLSKALKD-----IQSGALDINKAGILYGPQKTLHLLEA 154		
DB	63 PSQVRPD----GLRSGDGVPPRSIQDGTREGFGHSTSLKVPPLA-----RSLQISEEL 110		
QY	155 LPAGK-----PASFKNKTRFDHDSYSYKDSKETCAVLQKVALWARAQAE-RTEKSKL 205		
DB	111 LSRNQLSTASLPGFSGLQN-----HGQH-----LILSRASNAKPHYEFLSRMKF 156		
QY	206 N----LLETSEIKFPTASTYLHQLTKMKVTPQFKNESLQYETSNTVOLKTPQLRVSS 261		
DB	157 RGNALNSISDLPLAENS-----APFKMAHQTKQDKR-DMHSHP-VDLKLPQVRGMD 209		
QY	262 VKSQPDGSGGLDVMYQVSKTSSV-----EGSALQKLKNILPKQNKIEC--SGFVTHSSV 315		

Db	210	LSWESRTGD-----QYSSSLVMSQSTESALSCKLRAILPKQNRKSMLDAGP-----	256
QY	316	DSYFLHGLDLSPLCLNSKNGTVDGTSENTEDGLDKDSQPRKRGYRYDHEIMEEATA	375
Db	257	DSWGSDAE-----QSTSGQPYPTSDQEGD-----PGSKQPRKRGYRYQNSEILLEAIS	306
QY	376	MYMSGMSVSKAQGIYGVPHSTLEYKVKERSGTLTTPPKKLR-----PDTGL	424
Db	307	VMSGKMSVSKAQSIYGIPIHSTLEYKVKERLGLTKNPPKPKMKLMRSEGDVSV	360
RESULT 13			
Q8C9B1		PRELIMINARY;	PRT; 223 AA.
Q8C9B1			
AC	Q8C9B1;		
DT	01-MAR-2003	(TREMELrel. 23, Created)	
DT	01-MAR-2003	(TREMELrel. 23, Last sequence update)	
DT	01-MAR-2003	(TREMELrel. 23, Last annotation update)	
DE	Mus musculus	7 days neonate cerebellum cDNA, RIKEN full-length	
DE	enriched library, clone:A730006J22	product:hypothetical protein, full	
DE	insert sequence.		
OS	Mus musculus	(Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=Cerebellum;		
RX	MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;		
RA	Carninci P., Hayashizaki Y.;		
RT	"High-efficiency full-length cDNA cloning.";		
RL	Meth. Enzymol. 303:19-44(1999).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=Cerebellum;		
RX	MEDLINE=21085660; PubMed=11217851; DOI=10.1038/350555500;		
RA	RIKEN FANTOM Consortium;		
RT	"Functional annotation of a full-length mouse cDNA collection.";		
RL	Nature 409:685-690(2001).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=Cerebellum;		
RA	The FANTOM Consortium,		
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;		
RT	"Analysis of the mouse transcriptome based on functional annotation of		
RT	60,770 full-length cDNAs.";		
RL	Nature 420:563-573(2002).		
RN	[4]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=Cerebellum;		
RX	MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;		
RA	Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,		
RA	Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;		
RT	"Normalization and subtraction of cap-trapper-selected cDNAs to		
RT	prepare full-length cDNA libraries for rapid discovery of new genes.";		
RL	Genome Res. 10:1617-1630(2000).		
RN	[5]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=Cerebellum;		
RX	MEDLINE=2030913; PubMed=11076861; DOI=10.1101/gr.152600;		
RA	Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,		
RA	Konno H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itoh M.,		
RA	Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Kawaguchi K.,		
RA	Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,		
RA	Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,		
RA	Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,		
RA	Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;		
RT	"RIKEN integrated sequence analysis (RISA) system-384-format		
RT	sequencing pipeline with 384 multicapillary sequencer.";		
RL	Genome Res. 10:1757-1771(2000).		
RN	[6]		
RP	SEQUENCE FROM N.A.		

"Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes."; Genome Res. 10:1617-1630(2000).

[5]

SEQUENCE FROM N.A.

STRAIN=C57BL/6J; TISSUE=Cortex;
MEDLINE=30530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,
RA Konno H., Akhiyama J., Nishi K., Kiteunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,
RA Onoda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RA "RIKEN integrated sequence analysis (RISA) system-304-format
sequencing pipeline with 384 multicapillary sequencer.";
Genome Res. 10:1757-1771(2000).
[6]

SEQUENCE FROM N.A.
STRAIN=C57BL/6J; TISSUE=Cortex;
ADachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayaehida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Horii F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kohjima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Oheato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toyu T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
EMBL; AK043845; BAC31678.1; --
DR MGD; MGI:2651932; Mlr1
DR GO; GO:0005634; C:nucleus; IC.
DR GO; GO:0003702; FRNA polymerase II transcription factor acti. . ; IDA.
DR GO; GO:0006366; P:transcription from Pol II promoter; IDA.
KW Hypothetical protein.

SQ SEQUENCE 315 AA; 34830 MW; CD99B448E52A650C CRC64;

Query Match 19.7%; Score 444; DB 2; Length 315;
Best Local Similarity 80.4%; Pred. No. 7.8e-19;
Matches 90; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

QY 1 MKKMIQPALEYISKGGTKQENRNGSIGPSIVCKSIQMNAENSLOEQBGPLDTVNRM 60
Db 160 MKKMIQPALEYISKGGTKQENRNGSIGASLVCKSIQMNAENCLQEQBGPLDTVRT 219

QY 61 QEQTQQGQGVLDLTSTKTSIKSEESSICDPSSNSVAGRLHRNREDDYVERS 112
Db 220 QEQTQQGQGVLDLTSTKTSIKSEESSISDPSSNAVAGMLQMKTEKDVLDS 271

RESULT 15

Q8BRT8 PRELIMINARY; PRT; 315 AA.

CS Q8BRT8;
DC 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Mus musculus 10 days neonate cortex cDNA, RIKEN full-length enriched
DE library, clone:A830001M15 product:hypothetical protein, full insert
DE sequence.
GN Name=Mlr1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]

SEQUENCE FROM N.A.
STRAIN=C57BL/6J; TISSUE=Cortex;
MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.

